AL-TP-1992-0048

AD-A257 711

ARMSTRONG LABORATORY

AUTOMATION INFORMATION PROCESSING AND HIGH PERFORMANCE SKILLS: INDIVIDUAL DIFFERENCES AND MECHANISMS OF PERFORMANCE IMPROVEMENT IN SEARCH-DETECTION AND COMPLEX TASKS

Arthur D. Fisk
Kevin A. Hodge
Wendy A. Rogers
Mark D. Lee
Christopher K. Hertzog
Gregory M. Corso

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF PSYCHOLOGY
ATLANTA, GA 30332-0170

HUMAN RESOURCES DIRECTORATE
LOGISTICS RESEARCH DIVISION

SEPTEMBER 1992

INTERIM TECHNICAL PAPER FOR PERIOD NOV 1990 -DEC 1991

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation, or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

Beverly A. Gable

BEVERLY A. GABLE
Contract Monitor

BERTRAM W. CREAM, Chief
Logistics Research Division

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 1992	**3. REPORT TYPE AND DATES COVERED** Interim - Nov 1990 to Dec 1991

4. TITLE AND SUBTITLE
Automatic Information Processing and High Performance Skills: Individual Differences and Mechanisms of Performance Improvement in Search-Detection and Complex Tasks

5. FUNDING NUMBERS
C - F33615-88-C-0015
PE - 61101F
PR - ILIR
TA - 40
WU - 01

6. AUTHOR(S)
Arthur D. Fisk, Mark D. Lee
Kevin A. Hodge, Christopher K. Hertzog
Wendy A. Rogers, Gregory M. Corso

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Institute of Technology
School of Psychology
Atlanta, GA 30332-0170

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Logistics Research Division
Wright-Patterson AFB, OH 45433-6573

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL-TP-1992-0048

11. SUPPLEMENTARY NOTES
Armstrong Laboratory Technical Monitor: Beverly Gable, AL/HRGA, WPAFB OH, (513) 255-7773

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This document summarizes Phase 3 of the basic research effort investigating automatic processing theory and high-performance skills training. Research issues such as skill acquisition, skill retention, and part-task training are explored. The studies were conducted to examine: individual differences in performance improvement in memory, visual, and hybrid memory/visual search; effects of varying degrees of inconsistency on skilled visual search; development of optimal search strategies; and part-task training effects in learning and retaining complex task performance. The results of this work suggest further investigation of the principles for the application of automatic processing theory to training complex skills.

14. SUBJECT TERMS
Automaticity, Skill retention
Part-task training, Skill transfer
Skill acquisition

15. NUMBER OF PAGES 374

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500 Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

PREFACE

The work documented in this report was conducted under Air Force Human Resources Laboratory (AFHRL) Contract No. F33615-88-C-0015 with the University of Dayton Research Institute and was performed by the subcontractor Georgia Institute of Technology Research Institute. This work supports an integrated research program which is developing advanced part-task training techniques based on information processing theory. Beverley A. Gable served as the AFHRL/LRG, Wright-Patterson Air Force Base contract monitor.

I. OVERVIEW OF THE EXPERIMENTAL INVESTIGATION

This document details five series of experiments conducted to further extend automatic/controlled processing research to command and control, mission-specific training. These experiments build upon and extend earlier investigations reported by Fisk, Hodge, Lee, and Rogers (1990); and Fisk, Rogers, Lee, Hodge, and Whaley (1991). The research addresses training-program-relevant research that can be broadly categorized as (a) acquisition, (b) transfer, and (c) retention of high-performance skilled behavior.

This document describes experiments that examine issues related to (a) performance-ability relationships in visual search; (b) learning mechanisms in memory, visual, and hybrid memory/visual search; (c) effects of the introduction of varying levels of inconsistency on the use and maintenance of automatic processing in visual search; (d) effects of training environment on development of optimal search strategies in visual search, (e) component training for memory-dependent complex tasks, and (f) retention of rule-based processing and procedural knowledge. Because of the breadth of the issues examined, each series of experiments is presented in a separate section.

Section II presents the results from a very large-scale investigation utilizing 70 subjects with diverse ability levels. The experiment is important because it investigates the performance/ability relationships across multiple sessions of practice on pure visual search and during target/distractor role reversal after substantial consistently mapped (CM) practice. The relationships between cognitive and speed abilities and performance on CM and variably mapped (VM) visual search tasks were assessed across 6000 practice trials and 840 transfer trials. LISREL

techniques were used to assess the influence of general ability, fluid, and crystallized intelligence, working memory, perceptual speed, semantic memory access, and psychomotor speed abilities on search performance.

The results suggest that performance improvements in visual search can involve factors such as learning general and optimal search strategies, and developing automatic processing. However, the type of search (CM vs. VM) determines which factors are involved in performance improvements. Improvement in CM visual search is a function of all these factors. Improvement in VM visual search is a function of learning general and optimal search strategies. Convergent results from the normative reaction time data as well as ability/performance models for the practice and transfer sessions support these conclusions.

The third major section of this report describes a training and transfer experiment conducted to examine the relationship between the type of search processes used during training and what is learned during training. In this experiment, participants were trained in one of three CM search conditions: (a) pure memory search; (b) pure visual search; or (c) hybrid memory/visual search. After 6720 practice trials subjects transferred to a different search condition (or were not transferred and served in a "control" condition). For example, participants trained in the pure memory search condition transferred to either pure visual search or hybrid memory/visual search, or continued to perform the pure memory search condition. Such transfer conditions were also created for the other two training conditions.

The training phase of the experiment showed a striking difference between performance in conditions for which load was induced by memory set size versus display set size. These differences were especially striking early in

practice. Within the first practice session, performance in the pure memory search condition was 278 ms faster than in the pure visual search condition and 367 ms faster than in the hybrid memory/visual search condition. Such findings suggest that if transfer is not an issue (see below), task load can be better increased by increasing memory set size rather than visual set size (assuming no possibility for interaction effects with other tasks).

The transfer data clearly show a dissociation between the task structure used during training and subsequent ability to transfer to other types of search tasks. Subjects trained in pure visual search and hybrid memory/visual search were capable of transferring to any of the search conditions, including pure memory search. However, those trained in pure memory search demonstrated limited transfer to either pure visual or hybrid memory/visual search. Clearly, although across the training conditions subjects saw the same stimuli and made consistent responses to those stimuli, the type of learning seems driven by the type of task. Although these points have been raised previously (for reviews see Fisk and Rogers, 1991; Shiffrin, 1988), the empirical data have not been available within search/detection tasks to directly address these hypotheses.

From a practical perspective, the transfer data suggest that individuals who must perform consistent tasks that sometimes require pure memory search and sometimes require pure visual search of the same material should receive pure visual search practice. Further, the aforementioned situation may not require training under a memory search situation (even if one might ultimately be required to perform such a task) if visual search practice is provided. Similarly, at least within the constraints imposed by the present experimental design, if CM hybrid memory/visual

search may be required, pure CM visual search training could be sufficient and more easily implemented.

The experiments discussed in Section IV were conducted to examine the influence of varying degrees of consistency on the use and maintenance of automatic processing in visual search. In this experiment, subjects were first given CM training to develop visual search skills (6000 practice trials, 3000 per CM category). One CM category (the Adjusted-Consistent condition) was then transferred to either 100-, 67-, 50-, or 33-percent consistent search. (The other trained CM category, the Continuously-Consistent condition, remained consistently mapped throughout this phase of the experiment.) Following practice in the degree of consistency phase, subjects returned to 100-percent consistent search conditions.

The subjects in all groups developed skilled visual search during the Training Phase. Their performance improved during training with the performance-practice function fitting a general power function. Performance improvement, which is well-described by the "ubiquitous law" of practice (Newell and Rosenbloom, 1981), was one indication that automatic processing had developed by the end of 6000 practice trials. In the Degree of Consistency Phase, performance on reversal trials indicated substantial disruption -- another indication of automatic process development.

The disruption on the Adjusted-Consistent condition (the category that underwent the degree-of-consistency manipulation) target trials was a function of degree of consistency; more disruption occurred as consistency decreased. Performance on the category that remained consistent did not change except for an expected slowing due to changes in distractors. Hence, because the Adjusted-Consistent conditions were disrupted differentially as a

function of degree of consistency and the Continuously-Consistent condition showed minimal, uniform disruption, it can be concluded that the differential disruption of (or the differential need to inhibit) one automatic process does not necessarily differentially affect automatic processing on other tasks. In addition, the present findings also suggest that changes in task context will not necessarily disrupt automatic processes.

When transferring to the Retraining Phase, in which 100-percent consistency was restored, performance in the Adjusted-Consistent condition was better than performance in the New CM condition, regardless of previous degree of consistency. Hence, even the group whose Adjusted-Consistent condition was only 33-percent consistent retained some level of the automatic process developed in the Training Phase. Yet, more detailed analysis of the data indicated that performance was disrupted more in the Adjusted-Consistent condition than the Continuously-Consistent condition. This is evidence that some disruption of the automatic process, although minimal, occurred during the Degree of Consistency Phase.

The present findings have practical implications. First, it appears that inhibiting one automatic process will not dramatically affect a different automatic process if both processes are independent. This finding is important because it suggests that part-task training can be developed to retrain one automatic process without interfering with other related but independent automatic processes. The present data also suggest that individuals will retain well-learned, automatic processes despite inconsistencies encountered in the operational environment (at least within the limits presently tested). From an operational perspective this is both positive and negative.

Section V reports on an experiment conducted to examine the generality of the development of optimal search strategies. Subjects were trained in conditions of varying levels of consistency. The ratio of CM to VM trials was also manipulated to examine whether amount of practice was the primary factor leading to performance improvement in terms of optimal feature search. The pure CM condition followed a pattern expected for CM tasks for all groups of subjects. Although only 1200 trials of practice were provided, the Continuously-Consistent condition resulted in better final-level performance than the other conditions. Interestingly, the pure CM condition for the group that received the most inconsistency in the other conditions was slower after training (but not in the first session) than the other groups.

Final-level performance on the Adjusted-Consistent condition was as expected given the degree of consistency manipulation. Performance was fastest for Degree-Group 100, somewhat slower for Degree-Group 67, and slowest for Degree-Group 33. In fact, reaction time (RT) on the Adjusted-Consistent condition was about 100 ms slower for Degree-Group 33 than Degree-Group 100. Reversal performance was slower (nonsignificantly) than performance in the Supplemental-VM condition.

VM performance improved for all Degree-Groups: but the improvement did not reach the level of performance on the Continuously-Consistent condition for any Degree-Group. Improvement in VM did not follow the pattern expected, based on Fisher's (1986) feature overlap model. Search strategies appear to be developing for all groups; however, although Degree-Group 33 was the slowest (as predicted), Degree-Group 100 showed only a 23-ms faster RT than Degree-Group 33 in the VM condition. This is surprising because Degree-Group 100 received 3600 trials of VM (828 target trials, 207

target trials per VM category) while Degree-Group 33 received only 1200 VM trials (396 target trials, 99 target trials per VM category). Further problems for Fisher's theory arise because Degree-Group 67 had the fastest VM performance and that group received 3000 VM trials (750 target trials, approximately 187 target trials per VM category).

The ratio of CM to VM trials does not appear to be all the information needed to predict final-level performance in those training conditions. If such information were sufficient to predict performance, comparisons between the Continuously-Consistent condition and the VM condition should show the least difference between these conditions for Degree-Group 100, intermediate difference scores for Degree-Group 67, and the largest difference for Degree-Group 33. The differences between the Continuously-Consistent condition and VM condition were -88 ms, -66 ms, and -72 ms for Degree-Group 100, Degree-Group 67, and Degree-Group 33, respectively. These data show the value of understanding the total training environment when attempting to predict the effects of training manipulations. The data are also important because they place limits on theories of performance improvement based solely on search strategies.

Section VI provides the results of an experiment conducted in two phases, training and retention, using our complex dispatching task. The experiment was conducted to examine the benefits of part-task training of the memory components of the task and the effect of prior knowledge of the whole task on part-task training benefits.

The dispatching task is a conceptual analog of the tactical resource allocation required in real-world, battle-management tasks. This experimental sequence continues our use of complex tasks to evaluate the effects of instructional techniques on performance improvement and the

transferability of our major findings to even more complex, multi-component tasks. The task has several procedural components, requires substantial declarative knowledge, and is heavily rule-based. The task is conceptually simple -- the subject must choose the optimum "operator" for a given "delivery." However, this requires that the subject learn the rules associated with determining load level, load type, and delivery location characteristics. In addition, the subject must learn to associate 27 drivers with various "license classes" (license classification determines who can perform the mission).

The present task requires memory scanning (subjects must hold a self-derived list of potential drivers in memory). The number of potential drivers (and hence, memory load) is manipulated across trials, thereby providing data which converge on issues previously addressed with simpler laboratory memory search studies. Subjects must learn rules associated with performing the task; thus, rule-based learning (necessary for most complex skill-based tasks) can be assessed. Subjects must decide when and how to optimally access help screens (a decision component), and they must scan a display to locate the optimum driver (corresponding to standard visual search tasks).

In the first phase of the experiment, high-performance-skill development as a function of whole-task versus part-task training (and type of part-task training) was examined. The part-task training was designed to train declarative knowledge needed for whole-task performance. However, the actual task performed in part-task training was simple and contextually unrelated to the whole task. Four groups of subjects were trained. One group received whole-task practice throughout the experiment. The other three groups received part-task training. Two of these groups were told exactly how the to-be-learned material would be used in the

whole task. One of the part-task training groups (the Instructions Last group) was told only that the material learned in the part-task training would be used later in a much more complex task.

All aspects of performance improvement for subjects in the whole-task training groups followed a "power law" of practice (Newell & Rosenbloom, 1981). Early in practice, there were large individual differences in task performance. However, in line with other studies of skill acquisition (e.g., Ackerman, 1988; Fisk, McGee, & Giambra, 1988), these differences diminished with practice. Within the hours of practice, all subjects in the whole-task practice group increased accuracy (to ceiling), increased speed of decisions, reduced their use of help to very infrequent usage, and used the minimum number of keystrokes required.

The part-task training groups showed the most striking effects. The performance of subjects who received contextually relevant instructions regarding how the to-be-learned material would be ultimately used was strikingly better than the performance of those told only that the to-be-learned material would be used in a more complex task. The benefit of instructions relevant to the whole task was evident throughout part-task training.

Whole-task-trained subjects performed better than part-task-trained subjects at transfer, when all groups performed the whole task. However, the benefit of whole-task training was relatively small when compared to the benefit of a whole-task relevant briefing in conjunction with part-task training. In contrast, the part-task-training group that received no contextually relevant instructions performed worse than the other groups when transferred to the whole task. Subjective workload measures (NASA-TLX) were in agreement with the behavioral data.

The second phase of the experiment examined subjects' ability to perform the complex task 60 days subsequent to the final transfer session. The data indicated that although performance declined relative to the final training session performance, savings were impressive for all groups. However, the relative rankings of the groups were maintained across the retention interval.

Performance and retention characteristics followed the patterns expected from high-performance-skills development. Importantly, the data demonstrate the value of simplified part-task training for enhancing declarative knowledge needed to perform complex decision-making tasks. Just as important, the part-task-training data clearly show the need to provide instructions regarding ultimate use of to-be-learned material prior to providing part-task training.

The final section, Section VII, presents an augmentation of processing principles. These processing principles illustrate human performance guidelines that have been important in the development of "knowledge engineering" for understanding and developing training programs for complex, operational tasks. In addition, a summary of the lessons learned from the series of experiments presented in this report is provided.

Each section of this report is relatively self-contained. Therefore, the reader interested in specific issues need only read the relevant section(s).

II EXPERIMENTAL SERIES 1: AN INDIVIDUAL DIFFERENCES INVESTIGATION OF SKILL DEVELOPMENT IN VISUAL SEARCH

Introduction

There is a developing literature in cognitive psychology in which researchers are taking an individual-differences approach to understanding the processes involved in skill acquisition. Of particular interest are the specific underlying abilities which are important for skill development (Ackerman, 1984, 1986, 1988; Fleishman, 1972; Fleishman and Hempel, 1954; Kyllonen, Tirre, and Christal, 1991; Kyllonen and Woltz, 1989; Woltz, 1988). The logic of the approach is as follows. If a particular ability is important for successful performance of a task (or task component), individual differences in that ability should correspond to individual differences in task performance. Following Underwood's (1975) logic, if individual differences in ability do not correspond to individual differences in task performance, the original hypothesis of an ability/performance relationship is probably wrong (assuming a reliable and valid assessment of the ability and task performance). Utilizing such an individual-differences approach, it is possible to assess ability/performance relationships across practice and make inferences about the importance of different abilities for successful performance.

The purpose of the present study was to utilize an individual-differences approach to assess the ability correlates of performance in visual search tasks. Visual search is a crucial component of many skills trained within the Air Force; however, definitive descriptions of performance improvement in visual search continue to be elusive (Shiffrin, 1991).

The present assessment of ability/performance relationships in visual search is important for several reasons. First, Ackerman (e.g., 1988) has made general predictions about the relationships between abilities and search performance, but these predictions have not been tested for pure visual search. Second, previous ability/performance investigations have provided a limited amount of search practice (e.g., 720 trials in Ackerman, 1988) and assessed a limited number of abilities. These potential shortcomings are remedied in the present experiment in which ability/performance relationships in visual search were assessed as a function of extreme practice. Seventy young adults (ages 17-30) received extensive practice on the criterion task which was a semantic category visual search task (6000 total practice trials). Twenty ability tests were administered to measure the following factors: general intelligence, fluid intelligence, crystallized intelligence, working memory, perceptual speed, semantic memory access speed, and psychomotor speed.

Mechanisms of Improvement in Visual Search

In a visual search task, a single item is held in memory and compared to a visual display containing more than one item. The objective is to determine if (or which) one of the display items matches the item in memory (Atkinson, Holmgren, and Juola, 1969). The matching item is the "target" and the remaining items in the display are "distractors."

Visual search tasks have been a cornerstone in traditional attention research (Shiffrin, 1988) but there is not a theory of visual search performance improvement upon which everyone agrees. Practice-related changes in the attentional processes involved in the detection or localization of stimuli have been well-documented in the

visual domain. However, there has been a debate in the literature over the past decade or so about the mechanisms responsible for efficient visual search performance (Duncan and Humphreys, 1989; Fisher, 1984; Fisher and Tanner, in press; Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977; Treisman, 1982; Treisman and Gelade, 1980). These theories are briefly reviewed below to provide an overview of the mechanisms by which performance improves. The general conclusion that may be drawn from the literature is that several learning mechanisms are involved in visual search improvement. Both efficient search strategies and attention training contribute to changes in performance that occur with practice. This review is not exhaustive; rather, it highlights the elements of current theories which must be included in a general theory of visual search.

Several theories in the literature presumably explain (or at least describe) the practice effects that occur in visual search. Shiffrin and his colleagues (Schneider and Shiffrin, 1977; Shiffrin and Czerwinski, 1988; Shiffrin and Schneider, 1977) have proposed that pure visual search benefits most from an ability to differentiate (i.e., filter) targets from distractors (see also Dumais, 1979; Rogers, 1989). According to this theory, the critical variable for performance improvement is consistency. In a CM visual search task, particular items serve as targets <u>or</u> distractors but not both. As a result, every time a particular target item appears in a display it is attended and/or responded to. After many CM practice trials (sometimes thousands), an automatic response will be associated with the target item; that is, the CM target category will attract attention preferentially, relative to the other items in the display (Shiffrin, 1988). Theoretically, performance in such a CM-trained visual search task will be independent of visual load; that is, if an automatic response is associated with a stimulus,

attention-demanding search is not necessary. In a VM task, the same item might serve as a target on one trial (and be attended to) but a distractor on another trial (and be ignored). This inconsistency prevents the association of an automatic response with particular stimulus items.

Another theory of performance improvements in visual search has been proposed by Fisher (1982, 1984; Fisher and Tanner, in press). Fisher's model is based primarily on the concept of optimal, skilled search. Individuals learn not only which features to search for but also the optimal order in which to search for those features (i.e., a target feature sequence). Fisher's work has been instrumental in demonstrating that consistency of mapping targets to distractors is not the sole determinant of practice effects in visual search. Instead, according to Fisher, featural overlap and search strategies play a critical role in the development of efficient search performance. However, consistency at a featural level is also an important variable in his model.

Duncan and Humphrey's similarity-based theory (1989) and Treisman's feature integration theory (e.g., Treisman and Gelade, 1980) do not address practice effects, per se, but they do provide important information about the role of stimulus characteristics in visual search. According to the feature integration theory, visual search efficiency is a function of whether the search can be carried out pre-attentively (and in parallel) or if it is a serial process that requires a conjunction of features. The seriality of search may also be a function of the similarity of features such that even when a conjunction of features is not required, a high degree of similarity between targets and distractors may result in serial search (Treisman and Gormican, 1988).

According to Duncan and Humphreys (1989), performance in visual search is a function of the similarities and dissimilarities of potential targets and distractors. The most efficient search occurs when there is a high degree of dissimilarity between targets and distractors, along with a high degree of similarity among the distractor items.

These theories of performance and practice effects in visual search propose seemingly disparate mechanisms as the primary determinant of visual search performance. However, there is some evidence to support all the perspectives (Shiffrin, 1991). Visual search is a relatively complex task and performance improvements occur on many dimensions. Important variables for performance improvement include the similarity and featural overlap of targets and distractors, the development of efficient search strategies, and the development of an automatic response when CM targets are used.

A commonality among the theories involves the idea of consistency: whenever there is consistency there will be learning. The differentiating variable among the theories is whether consistency must be at the featural level (Duncan and Humphreys, 1989; Fisher, 1984; Fisher and Tanner, in press) or at the more global, target-to-distractor level (Shiffrin and Schneider, 1977)--an integrative perspective can incorporate both levels. Consistency at the featural level allows for the development of optimal search strategies. Simultaneously, or subsequently, consistency of targets and distractors allows for the development of an automatic response. Fisher and Fisk (1991) suggest an alternative model of learning in search/detection tasks. According to their model, VM search practice will, at best, allow for performance improvement due to learning an "optimal search strategy." CM search practice allows for performance improvement due to learning an optimal search

strategy as well as developing an automatic response. This alternative account of the development of search-related skills seems necessary to incorporate the critical elements of the above-mentioned theories (most notably, strategic search and attention training). Some data support such a view of improvement in visual search (see Fisk, Rogers, and Lee, 1991; Rogers and Fisk, in press); however, converging evidence is needed.

Although in a different instantiation, similar ideas have been proposed by Rabbitt (e.g., Rabbitt, Cumming, and Vyas, 1979) to account for performance improvements in tasks with a visual search component. Rabbitt states that individuals learn to use specific sets of cues to optimally discriminate between relevant and irrelevant information. Furthermore, according to Rabbitt, practice results in the symbol-specific learning being superseded by other processes. Although Rabbitt does not specify what the additional processes might be, his basic ideas are consistent with the theory that general search strategies are superseded by automatic response development.

In the present experiment, an individual-differences approach was utilized to provide convergent evidence about the processes critical to performance improvement in visual search. To paraphrase Fleishman (1972), individual differences were exploited to gain insight about the processes required to perform a particular task. The experiment was not designed to provide the critical distinction between the aforementioned theories of visual search. In fact, if search strategies and attention training combine to yield the most efficient search performance, then an integrative theory (such as the framework proposed by Fisher and Fisk, 1991) is most appropriate, rather than one particular theory. Instead, the goal was to specify the abilities which are important

for performance improvements in CM and VM visual search. Comparisons of the ability/performance relationships for CM and VM practice conditions will aid in the understanding of the processes that are important for improvement in the two tasks.

Overview of Criterion Task: Semantic Category Search

Fisk and Schneider (1983; Schneider and Fisk, 1984) demonstrated that the basic characteristics of CM and VM performance proposed by Shiffrin and Schneider (1977; Schneider and Shiffrin, 1977) also apply to more complex search tasks in which the memory set consists of a category label (e.g., Fruit), and the target and distractor items in the display consist of category exemplars (e.g., Apple, Dog, etc.). The task is to determine if a word in the display is a member of the category in the memory set. This type of task, semantic category visual search, is used in the present experiment. This task was chosen for several reasons: it is a representative visual search task (i.e., performance characteristics are well-replicated in the literature); it is complex enough to show substantial performance improvements as a function of practice; and there has been some work with young adults on the relationships between abilities and category search performance (e.g., Ackerman, 1988).

Semantic category visual search can be described as a function of two stages. First, the relationship of an exemplar to its higher-order category is determined (i.e., an apple is a fruit). Second, the relationship between the category and the required response is determined. That is, if the category matches the memory-set category then a positive response is required; otherwise, no response or a negative response is required.

In both CM and VM category visual search tasks, the exemplar-to-category stage is consistent (i.e., an apple is always a fruit). Any learning that occurs in this stage will involve strengthening the category-to-exemplar link within the experimental context. This strengthening, which can occur in both CM and VM tasks, will be a function of the number of words in each category and the number of categories in each condition.

The category-to-response link (Stage 2) will be consistent only for the CM task in which categories are either targets or distractors but not both. In this condition, a particular category is always associated with the same response (i.e., targets are always attended to and distractors are always ignored). Strengthening the category-to-response link can result in the development of an automatic response to the consistent target category such that a serial search through the display is no longer necessary.

Recall that, by definition, the same category may serve as both a target and a distractor across trials in the VM task. Thus, on some trials, a particular category will be responded to (when it is the target) but on other trials that same category will not be responded to (when it is a distractor). Consequently, automatic response development is not possible in a VM task.

To summarize, both the CM and the VM tasks may be classified as learning tasks, albeit to different degrees. In both tasks, subjects will learn search strategies, response mappings, and so on. But, by design, an automatic response can only develop in the CM condition (because this task is completely consistent from stimulus to response). Thus, CM practice provides a situation in which automatic response development is possible, whereas VM conditions inhibit such development.

General search strategies can be assessed through transfer to new categories. If only general search strategies are learned, transfer to new target/distractor pairings should be quite good and equivalent to final-level CM performance. However, if final-level search performance is a function of an optimal search strategy that is stimulus-specific (i.e., category-specific), then New CM performance should be worse than final-level CM performance.

Automatic response development can be separated from an optimal search strategy in a different type of transfer condition. Shiffrin and Schneider (1977, Experiment 1) demonstrated that an inference can be made about automatic response development through the "reversal" of the CM targets and distractors. They trained CM targets and distractors, then reversed the roles of both within the same condition (i.e., the previous CM targets became distractors and the previous CM distractors became the targets). Shiffrin and Schneider found that performance in this full reversal condition was actually worse than asymptotic VM performance. Subjects required nearly three times as much practice to reach the level of CM performance prior to reversal. The disruption in performance at reversal is consistent with the theory that an automatic response was associated with the previous CM target. If CM performance is driven by an automatic response, then CM Reversal performance should be worse than New CM performance (due to the disruptive effect of an automatic response that is no longer compatible with the task). However, if performance in the CM Reversal condition is equivalent to performance in the New CM condition and both are worse than final-level CM performance, then performance is likely driven by an optimal search strategy which is stimulus-specific but does not involve an automatic response.

In the present experiment, subjects received 3000 CM practice trials and 3000 VM practice trials on a semantic category visual search task. Following this extensive practice, their performance was assessed in two transfer conditions. A New CM condition was created by pairing two of the VM categories into a consistent condition (thus the words were not completely novel). This New CM condition allowed an assessment of general search skills that are stimulus-independent. Note that the New CM condition did not require new exemplar-to-category strengthening because the stimuli were familiar from the VM task. The second transfer condition was a CM Reversal; that is, the roles of the trained CM targets and distractors were reversed. Disruption in this condition provides an index of the automatic response developed for the consistent target (Dumais, 1979; Rogers, 1989; Shiffrin and Schneider, 1977).

Performance Predictions

Subjects were expected to show performance improvements (i.e., faster reaction times) in both the CM and VM versions of the task. Under both practice conditions, subjects benefit from learning general and/or optimal search strategies, where to look on the screen, and the location of the response keys, as well as strengthening the exemplar-to-category links for the categories used in the experiment. More improvement in terms of comparison slope estimates (increase in reaction time corresponding to an increase in display size) was predicted for the CM version of the task because only in this condition was automatic response development possible.

Given the hypothesis that young adults develop an automatic response to highly trained CM categories, the transfer conditions were expected to cause a differential disruption in performance. In the New CM condition, an automatic response is not attached to the new category.

Thus, performance is worse than final-level CM performance, but should be better than CM Reversal performance. In the CM Reversal condition, the automatic response to the previously trained CM target category (which is now serving as the distractor category) severely disrupts performance. Subjects have difficulty ignoring the previously trained CM target; performance in the CM Reversal condition is worse than final-level CM performance and worse than performance in the New CM condition. Not only are subjects searching for a new category in the CM Reversal condition, but they must also inhibit the automatic response to the previous CM category (now serving as a distractor) which was developed in practice.

Analysis of Ability/Performance Relationships

The general idea that ability requirements will vary at different stages of practice was stated at least as early as 1899 by Bryan and Harter and has been echoed over the years by a variety of researchers (e.g., Ackerman, 1988; Adams, 1987; Carroll, 1988; Corballis, 1965; Ferguson, 1954, 1956; Fleishman, 1972, 1975; Fleishman and Hempel, 1955; Guilford, 1967, 1985; Humphreys, 1960; Kyllonen and Woltz, 1989; Labouvie-Vief, Frohring, Baltes, and Goulet, 1973; Peterson and Barlow, 1928; Sternberg, 1985; Woodrow, 1946). There has been some controversy, however, over whether individuals change (i.e., the individuals' abilities change) as a function of practice on a task or whether the task changes. Henry and Hulin (1987) cite Adams (1957), Corballis (1965), and Alvares and Hulin (1972) as providing support for the changing-subjects model. According to this view, the specific abilities required for a task remain constant but the individual's level of ability changes and improves. According to a changing-task model, the type of abilities required for a task changes as a function of time-on-task or practice (Fleishman and Hempel, 1955; Ackerman, 1989).

According to the results reported by Ackerman (e.g., 1988), Fleishman and Hempel (1955), and Labouvie-Vief et al. (1973) and the theories proposed by Ferguson (1954, 1956) and Guilford (1967), as skill acquisition proceeds, the nature of the task changes and requires different abilities for successful performance. However, it remains to be determined if these changes in task requirements also influence ability levels.

Fleishman conducted extensive research to investigate the relationship between abilities and improvements with practice on motor tasks (Fleishman, 1972; Fleishman and Hempel, 1955). He was interested in being able to predict the performance of individuals after they experienced a long training program. He found, however, that initial performance on complex tasks was a poor predictor of final-level performance. Fleishman (1972, 1975) proposed three basic principles to explain ability/performance correlational changes with practice (1) broad cognitive abilities determine initial task performance; (2) perceptual motor abilities increasingly determine performance later in practice; and (3) some new, task-specific ability develops with practice which differs from both the cognitive and perceptual motor abilities. Unfortunately, Fleishman's results must be questioned because he used inappropriate data-analysis methods (Corballis, 1965; Humphreys, 1960). Fleishman factor-analyzed his practice data along with the data from the ability measures. Factor analysis of the simplex pattern results in artifactual effects from this analytic approach. Humphreys (1960) and Corballis (1965) note that a practice matrix cannot be included in a factor analysis matrix, but factor structure of the criterion tests should be independently determined.

Based in part on a reanalysis of Fleishman's data, Ackerman (1987) has recently proposed an integrative theory

of individual differences in learning. Ackerman's basic premise is that the abilities which determine performance differ as a function of skill acquisition. Importantly, the ability/performance relationships can be predicted given knowledge of skill level and are not task-specific as proposed by Fleishman.[1] To briefly review, Ackerman defines general ability according to Humphreys' (1979) hierarchical model, and perceptual speed and psychomotor speed are defined as first-order factors that are orthogonal to the higher-order general ability factor. Ackerman's conception of skill acquisition is consistent with general theories in which skill is acquired in stages ranging from initial, novice performance to intermediate to final-level, skilled performance (e.g., Anderson, 1982, 1983; Fitts, 1964; Fitts and Posner, 1967; Shiffrin and Schneider, 1977).

The ability/performance predictions from Ackerman's theory are illustrated in Figure 1. Novice-level performance corresponds to demands on general and content abilities.[2] Individual differences in these abilities are related to individual differences in amount or efficiency of attentional resources. Consequently, initial performance will be at least moderately correlated with general ability, and, to some degree, with task-relevant content abilities such as verbal or spatial ability (the degree of association will be dependent on task complexity).

Intermediate stages of skill acquisition correspond to demands on perceptual speed abilities. Perceptual speed ability involves "speed in finding figures, making

[1]The description of Ackerman's model provided here has been culled from his various papers on the topic (Ackerman, 1984, 1986, 1987, 1988, 1989; Ackerman and Schneider, 1985). However, an effort has been made to maintain the spirit of the most recent descriptions because they obviously represent the result of the transitions and developments of his general theory.
[2]Content abilities correspond to group factors such as verbal or spatial ability.

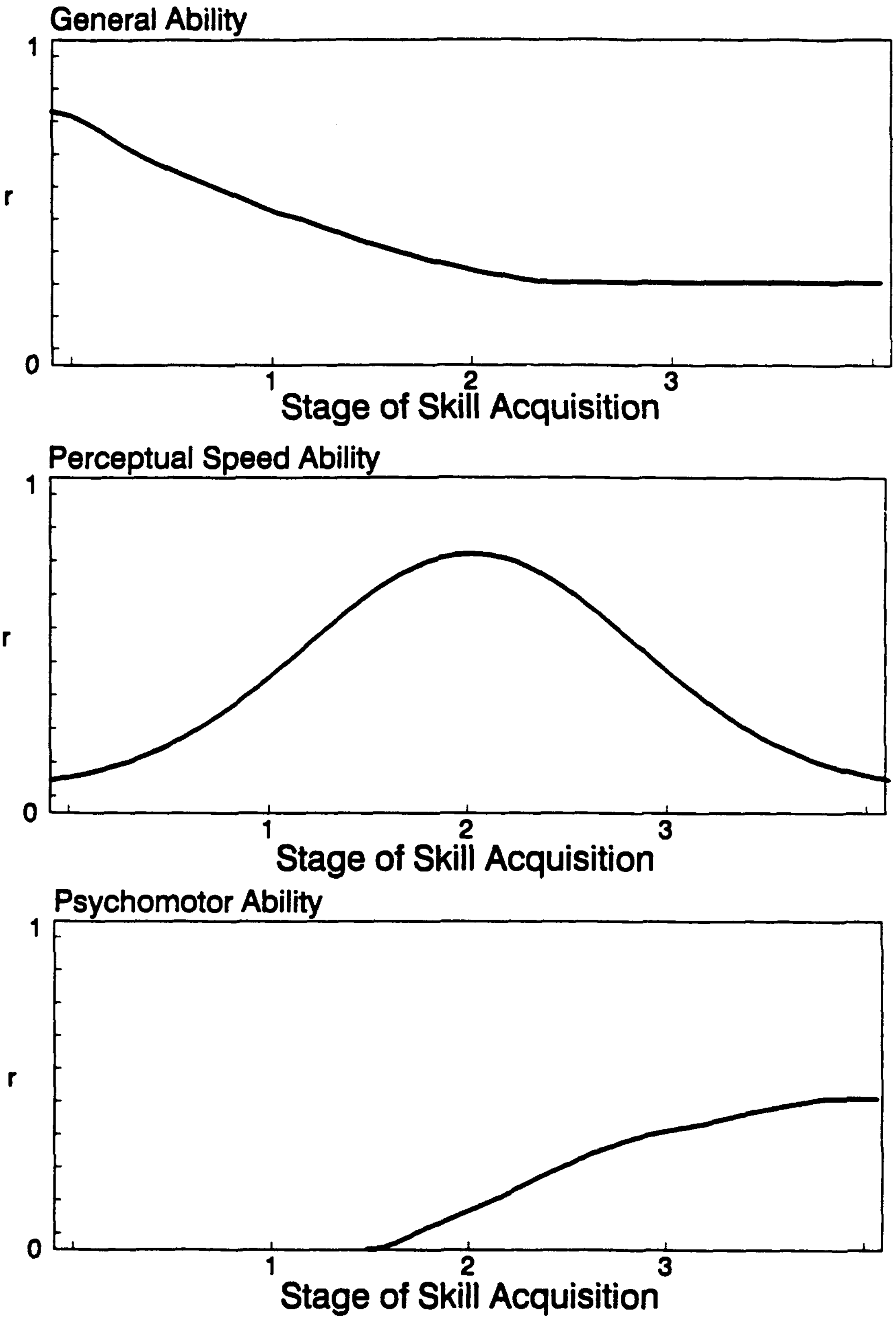


Figure 1. Schematic Representation of Ackerman's Theory (1988)

comparisons, and carrying out other very simple tasks involving visual perception" (Ekstrom, French, and Harman, 1979, p. 29). These processes may also be described as the generation and compilation of simple production systems (Anderson, 1982, 1983). According to Ackerman, the relationship between perceptual speed and practice is an inverted U-shaped function. Early in practice, production rules are still being tested; after some practice (the amount necessary is probably task dependent), these productions are being tuned--hence there is a greater relationship to perceptual speed ability. Finally, this tuning process asymptotes as skill is attained and the importance of perceptual speed ability is attenuated.

The final level of skill acquisition corresponds to predominantly noncognitive psychomotor abilities. Psychomotor speed represents processing speed which is independent of information processing; it is denotative of *psychophysical* limitations. Individual differences in psychomotor ability represent differences in "running-off" fine-tuned productions and determine individual differences in skilled performance.

Ackerman (1984, 1986, 1988; Ackerman and Schneider, 1985) has conducted an extensive series of experiments designed to assess ability/performance relationships. Some of these experiments utilized semantic category search as the criterion task. Only the predicted relationships with general ability (G) and perceptual speed (PS) have been tested with these tasks (psychomotor ability measures were not included in the test battery), but the data generally supported the theory. To illustrate, in a CM search task for which the memory set and display set both consisted of three words (Ackerman, 1988, Experiment 2), initial-level performance correlated .37 with G (as determined with the Dwyer [1937] extension procedure) whereas final-level

performance correlated .09. PS correlated .17 with initial performance, rose to .35 after practice on the task, and correlated .25 with final-level performance. In a VM task for which the memory set consisted of two words and the display set consisted of three words (Ackerman, 1988, Experiment 4), the pattern of ability/performance relationships was similar, albeit the size of the correlations differed from the CM task. That is, there was a general decrease in the correlation of performance with G (.38 to .32) and an increase with PS (.15 to .35).[3]

It is important to note that in Ackerman's search tasks, memory load was manipulated and visual load was held constant (usually at three items). Thus, the tasks are hybrid memory/visual search tasks in that both memory load and visual load were greater than one. Recent research supports the idea that there are distinctions in the processes involved in memory and visual search (e.g., Czerwinski, 1988; Fisk and Rogers, 1991; Flach, 1986). Consequently, ability/performance relationships might also differ as a function of whether the task is primarily a memory search task or a visual search task. The present analysis focuses on the ability performance relationships in a "pure" visual search task. That is, memory load is always one item and visual load is varied (and greater than one). The results of this experiment will provide information about whether the ability/performance relationships differ for visual search relative to hybrid memory/visual search (through comparisons with Ackerman's results). Moreover, the present investigation assesses ability/performance relationships across a more extensive range of practice (3000 trials per condition compared to 720 trials of practice in Ackerman [1988]).

[3]These correlations are all approximations taken from Figures 9 and 13 of Ackerman (1988).

Ability Factors

Based on ability/performance models in the literature, the following ability factors were chosen for the investigation of ability/performance relationships across CM and VM practice conditions. Abilities are considered to be relatively stable characteristics of individuals that determine their performance level (Carroll, 1983).

General Ability. According to Fleishman (1972) and Ackerman (1988), general cognitive abilities (i.e., G) are critical for successful performance of novel tasks. A higher-order G factor will be defined and its influence on search performance will be assessed.

Fluid Intelligence and Crystallized Intelligence. Cattell (1963) proposed that there are two "general abilities": fluid intelligence (Gf) and crystallized intelligence (Gc). Fluid intelligence is described as the capacity to perceive relations and educe correlates critical for adapting to new situations. Gc denotes our storehouse of knowledge--the stored results of prior Gf applications. Separate Gf and Gc factors will be defined to determine if, in addition to the higher-order G factor, either Gf or Gc is predictive of initial-level performance.

Working Memory. Working memory (WM), as described by Baddeley (1986), involves the simultaneous processing and storage of information. The concept of WM is central to many skill acquisition models (e.g., Anderson, 1983; Carlson, Sullivan, and Schneider, 1989; Schneider and Detweiler, 1987). For example, Anderson (1983) suggested that WM capacity should affect the initial declarative stage of skill development because a relatively large amount of knowledge about the skill must be acquired and interpreted. Once productions are established, the importance of WM may be reduced. In fact, Woltz (1988) has demonstrated that,

for young adults, WM shows a significant correlation with early performance on a learning task.

Perceptual Speed. Perceptual speed may be defined as speed in making visual comparisons. Recall that PS ability is postulated to be the performance-limiting factor during intermediate stages of skill acquisition (e.g., Ackerman, 1988, 1989; Fleishman, 1972, 1975). It is predicted that the influence of PS on performance will be low for initial-level performance but will show a pattern of increasing then decreasing effects on performance.

Semantic Memory Access. Semantic memory access (SMA) denotes the speed of retrieving information from long-term memory. In a recent analysis of ability/performance relations during the acquisition of procedural skills, Woltz (1988) suggested that automatic activation speed (his term for what is herein referred to as SMA) would be the limiting factor during the later phases of skill acquisition (see also Kyllonen and Woltz, 1989). However, in the current study, given the nature of the semantic category search task, it is also possible that SMA may show an influence on performance earlier in the course of practice, especially for the VM task. In other words, SMA may be related to the pre-experimental exemplar-to-category link and individual differences in SMA may correspond to individual differences on the task.

Psychomotor Speed. Psychomotor speed (PM) is the speed of simple responding in the absence of information processing requirements. Ackerman (1988, 1989) suggests that final-level performance is limited by psychomotor skills. These abilities differentiate individuals during this phase because all other task components have been automatized. However, relationships between psychomotor abilities and search performance have not been assessed previously.

General Predictions: Ability/Performance Relationships

Although the analytical approach to structural modelling was exploratory (see below), general descriptive predictions were made on the basis of previous investigations of ability/performance relationships (e.g., Ackerman, 1988; Fleishman, 1972). Initial search performance for both CM and VM was expected to be predicted by G, and the relationship between G and performance was expected to be reduced as a function of task practice (more so for CM than VM).

Given the nature of the semantic category search task, it was expected that SMA would also predict initial performance, perhaps to a greater extent for VM relative to CM because there were more categories in the VM condition. PS ability was predicted to be related to performance after some amount of practice, but the specific amount was unknown. The relationship of PS to performance was expected to be minimal by the end of practice.

Psychomotor speed was predicted to relate to final-level performance in the CM condition (i.e., skilled performance). This prediction was not made for the VM condition. The relationship between PM and performance is, according to Ackerman (1988), an indicator of automatic response development. An automatic response cannot develop for VM practice; thus, the PM/performance relationship was not predicted for that situation.

The relationships of the remaining abilities to performance could not be specifically predicted on the basis of previous work. However, it was assumed that if Gf, Gc, and WM were predictive of individual differences in search performance, they would be predictive of initial, novice-level performance.

Ability/performance relationships for the transfer session were expected to reflect the degree to which the transfer conditions were performed differently than the CM condition (at the end of practice). For example, if G were predictive of transfer performance, it would suggest similarities between novice-level and transfer performance. However, if final-level CM performance were highly predictive of transfer performance, it would suggest that, although all subjects might be slowed at transfer, their rank-orderings would not change. Hence, the inference would be that similar abilities predicted CM and transfer performance.

Structural Equation Modelling

A causal modelling approach was used to directly test the influence of abilities on performance. The first step in the modelling process was to determine which abilities were influential for initial-level performance. The second step was to fit the data with a first-order autoregressive process (Joreskog, 1970; Joreskog and Sorbom, 1977). In other words, once initial-level performance has been predicted, is later performance best predicted from earlier performance? Next, an exploratory approach was used to determine if abilities had any direct influence on later performance.

Although certain ability/performance relationships were predicted, the analytical approach was exploratory in nature. The entire model of ability/performance relationships could not be specified a priori; consequently, a specification search was conducted. Hertzog (1990) describes this approach as appropriate when the goal of the research is descriptive and the purpose is explorative. A restricted-factor analysis technique was utilized in which the initial descriptive hypotheses were tested and

subsequent model fitting was done on the basis of the data (Alwin, 1988; Hertzog, 1985).

Using the structural equation modelling approach allows more precise measures of partial covariances. This is particularly useful in the present context because the relationships between specific abilities and performance can be assessed after controlling for the influence of general, higher-order ability.

Another advantage of using the structural modelling approach is that ability/performance relationships are assessed at the latent-variable level (Hertzog, 1985). In essence, only the true abilities and true performance characteristics are compared because the latent factors in the model are attenuated for measurement error. Consequently, it is more likely that the relationships will be independent of particular measurement errors in the current sample. Thus, they will be replicable across subject samples. Furthermore, because of the attenuation for measurement error, the correlation between latent variables is generally higher than that between manifest variables (Everitt, 1984).

Summary of Experimental Approach

In the present experiment, the focus was on investigating ability/performance relationships across extensive practice. The goal was to determine which abilities best predict novice-level performance and whether the specific abilities, or degree of influence, change across practice. The use of structural modelling allowed a more precise analysis of the causal relationships between abilities and performance. The design of the experiment allowed an assessment of the ability/performance relationships for CM, VM, and transfer performance. Such an approach provides multiple opportunities to evaluate the

underlying abilities which are related to visual search performance and how the influences of particular abilities change (or do not change) across practice.

METHOD

Subjects

Seventy subjects were recruited from a large metropolitan community through newspaper advertisements and from a southeastern university campus. The demographic characteristics of the subjects are presented in Table 1. The subjects' corrected or uncorrected visual acuity was at least 20/40 for distance and 20/40 for near (equivalent to reading magazine print at normal reading distance). Students received course credit or $5.00 per hour for their participation; the remaining subjects received $5.00 per hour plus parking expenses. During a pre-experiment telephone interview, subjects' medications were recorded and rated in terms of the severity of effects on attention (Giambra and Quilter, 1988). Those subjects taking more than two drugs which have more than minimal effects on attention were not included in the experiment.

Ability Tests

Twenty ability tests were administered, including paper and pencil measures as well as computer tasks. All testing was done with groups of six or fewer subjects. Ability tests were chosen to enable the estimation of the following constructs: general intelligence, fluid and crystallized intelligence, PS, PM, WM, and SMA. Each construct of interest had at least three marker tests. All the tests in the battery are represented in Figure 2 along with the construct they are presumed to measure. The sources for the tests are presented in Table 2. The individual tests are described below.

Table 1. Subject Characteristics

Number	70	
Male/Female	46/24	
	Mean	SD
Age	20.83	.03
Self-rated Health[a]	1.46	.56
Formal Education[b]	13.48	1.72

[a] 1 - excellent, 5 - poor.

[b] 12 - high school, 16 - college, etc.

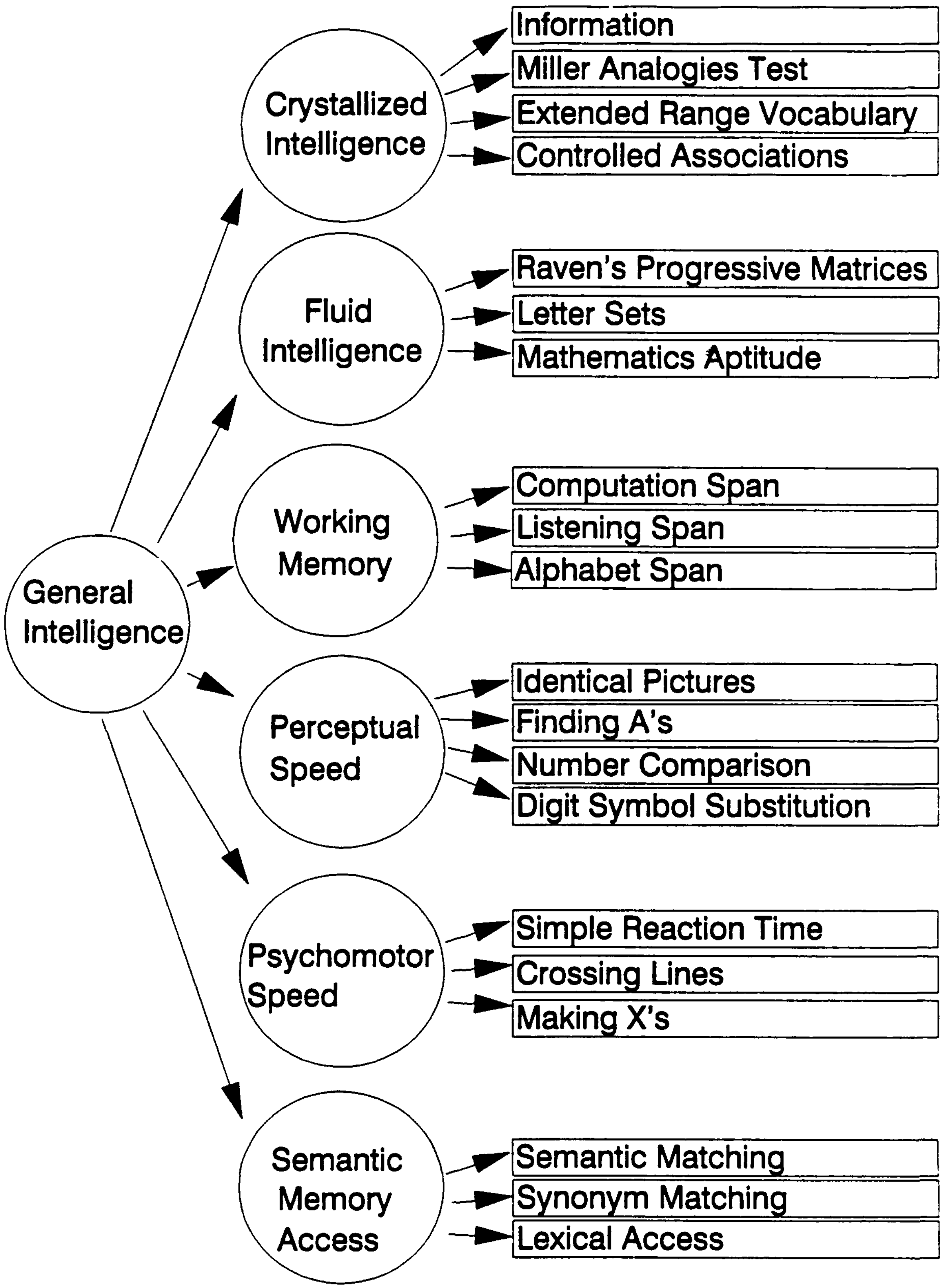


Figure 2. Proposed Factor Structure and Ability Tests

Table 1. Subject Characteristics

Number	70	
Male/Female	46/24	
	Mean	SD
Age	20.83	.03
Self-rated Health[a]	1.46	.56
Formal Education[b]	13.48	1.72

[a] 1 - excellent, 5 - poor.

[b] 12 - high school, 16 - college, etc.

Raven's Progressive Matrices (Raven, Court, and Raven, 1977). Subjects were presented with a 3 x 3 matrix of items (with the lower right item missing). Their task was to determine the rule by which the items changed across the rows and down the columns. Subjects were asked to choose the best option (out of eight) to replace the missing item. Part 1 consisted of 12 items and a five-minute time limit. This section was considered practice, as suggested by Raven et al. (1977), to familiarize subjects with the test and presentation method. Part 2 consisted of 36 items and a 40-minute time limit. The score was the total number correct for Part 2.

Letter Sets (Ekstrom, French, Harman, and Dermen, 1976). Subjects were presented with five sets of four letters. Their task was to find the rule that related four of the letter sets to one another and cross out the letter set that did not fit (i.e., did not follow the rule). Two parts of the test were administered. Each part consisted of 15 items and a seven-minute time limit. The score was the total number correct.

Mathematics Aptitude (General Reasoning) (Ekstrom et al., 1976). Subjects were required to solve short problems that required arithmetic or very simple algebraic concepts. The test was in multiple choice format and offered five possible answers. Two parts of the test were administered. Each part consisted of 15 items and a ten-minute time limit. The score was the total number correct.

Miller Analogies Test (Bader and Burt, 1983). Subjects were presented with a verbal proportion of the form A:B :: C:D with one of the terms missing (as in the Miller Analogies Test). There were four options for the missing term. Subjects were required to choose the option which best completed the analogy. There were 50 items and a 25-minute time limit. The score was the total number correct.

Extended Range Vocabulary (Ekstrom et al., 1976). This test, in multiple choice format, presented subjects with a probe word and five words as possible answers. Their task was to mark the word synonymous to the probe word. Two 24-item parts were administered, each with a six-minute time limit. The score was the total number correct.

Controlled Associations (Ekstrom et al., 1976). Subjects were required to write as many synonyms as possible for a given word. Two parts were administered; each consisted of four words and a six-minute time limit. The score was the total number correct.

Information (Wechsler, 1981). Short-answer, general-information questions were read orally to the subjects who were required to write down the correct answer. The test had a total of 29 questions and no time limit. The score was the total number correct.

Listening Span (Salthouse and Babcock, 1991). Subjects were required to answer questions about simple sentences, presented orally, marking the correct responses on an answer sheet. At the same time, the subjects were asked to remember the last word of each sentence. One to seven sentences were presented prior to the recall portion of the task. Following the sentences, the subjects were required to recall the final word of each sentence in the order in which they were presented. Three trials were presented at each level prior to progression to the next level (i.e., three trials with one sentence, then three trials with two sentences, and so on up to seven sentences). Span score was the total number of words recalled for trials that were recalled perfectly (absolute span, LaPointe and Engle, 1990).

Computation Span (Salthouse and Babcock, 1991). Subjects were required to solve simple arithmetic problems,

presented orally, by marking the correct responses on an answer sheet. At the same time, the subjects were asked to remember the last number of each problem. One to seven arithmetic problems were presented prior to the recall portion of the task. Following the arithmetic problems, the subjects were required to recall the final number of each problem in the order in which they were presented. Three trials were presented at each level prior to progression to the next level (i.e., three trials with one problem, then three trials with two problems, and so on up to seven problems). Span score was the total of the number recalled for trials that were recalled perfectly (absolute span, LaPointe and Engle, 1990).

Alphabet Span (Craik, 1986). Two to nine words were presented orally to the subjects. Their task was to recall the words in alphabetical order. For example, if presented with dog, cat, boy, the subject would recall boy, cat, dog. Subjects were required to do the alphabetizing "in their heads," then write the words in alphabetical order. Three trials were presented at each level prior to progression to the next level (i.e., three trials with two words, then three trials with three words, and so on up to nine words). Span score was the total number of words recalled for trials that were recalled perfectly (absolute span, LaPointe and Engle, 1990).

Lexical Access (Locally Developed - Shaw and Rypma). Subjects were presented with a letter string (e.g., bisp) on a computer monitor. Their task was to quickly and accurately decide if the letter string was a valid English word. Following 20 practice trials, there were 100 trials on the task (50 words and 50 nonwords, randomly intermixed). The score was mean correct-trial RT for the "word" trials. Trials for which RT was below 100 ms or above 4000 ms were not included.

Semantic Matching (Hertzog, Raskind, and Cannon, 1986). In this task, two words were presented simultaneously on a computer monitor. The subject's task was to decide as quickly and accurately as possibly whether the two words were members of the same semantic category (e.g., elm, poplar) or from two different categories (e.g., elm, shirt).[4] Prior to the task, the subjects read the list of categories and words to be used. Following eight practice trials, there were 96 trials on the task (48 same and 48 different, randomly intermixed). The score was the mean correct-trial RT for the "same" trials. Trials for which RT was below 100 ms or above 4000 ms were not included.

Synonym Matching (Hertzog et al., 1986). In this task, subjects were presented two words simultaneously on a computer monitor. Their task was to decide as quickly and accurately as possibly whether the two words had a similar meaning (e.g., price, cost) or different meanings (e.g., sofa, horn). Following eight practice trials, there were 96 trials on the task (48 same and 48 different, randomly intermixed). The score was mean correct-trial RT for the "same" trials. Trials for which RT was below 100 ms or above 4000 ms were not included.

Finding A's (Ekstrom et al., 1976). Subjects were presented with lists containing 41 words. Their task was to mark the words which contained the letter "a." Two parts of the test were administered. Each part consisted of 20 lists and a two-minute time limit. The score was the total number of words marked correctly.

Digit Symbol Substitution (Wechsler, 1981). Subjects were presented with a key containing the digits one through

---

[4]The categories used in this task (natural earth formations, articles of clothing, fish, parts of a building, vegetables, and trees) did not overlap with the categories used in the semantic category visual search task.

nine and a symbol associated with each digit. Below the key were four rows of 25 digits. The task was to fill in the blanks below the rows of digits with the appropriate corresponding symbols. There were 100 items and a 1.5-minute time limit. The score was the total number correct.

Number Comparison (Ekstrom et al., 1976). Subjects were required to compare two numbers, determine if they were the same or different, and mark only those pairs which contained different numbers. The numbers ranged from three to 13 digits in length. Two parts of the test were administered. Each part consisted of 48 items and a 1.5-minute time limit. The score was the total number marked correctly.

Identical Pictures (Ekstrom et al., 1976). In this multiple choice test, subjects were presented with a probe picture and simultaneously given five other pictures as possible responses. Their task was to mark the picture that was identical to the probe picture. Two parts were administered. Each part consisted of 48 items and a 1.5-minute time limit. The score was the total number correct.

Making X's (Locally Developed - Rogers). Subjects were presented with rows of squares (.25 in. x .25 in.). Their task was to work from left to right, one row at a time, and draw an X in each square as quickly as possible. Two parts of the task were administered. Each part contained 168 boxes (14 rows of 12) and had a one-minute time limit. The score was the total number of X's made.

Crossing Lines (Botwinick and Storandt, 1974). Subjects were presented with 12 rows of eight short (1/4") lines. Their task was to work from left to right, one row at a time, and draw a slash through each line as quickly as possible. Two parts were administered, each with a

twenty-second time limit. The score was the total number of lines marked.

Simple Reaction Time (Locally Developed - Lee). In this computer-controlled task, the digit "1" was presented in the center of the computer screen and subjects were required to respond as quickly as possible by pressing the "1" key on the number keypad. The stimulus was preceded by an asterisk (*) in the center of the screen followed by random foreperiods of 800, 900, 1000, 1100, or 1200 ms. The variable times introduced time uncertainty into the task. This task was administered in three parts, each with 20 trials. The first ten trials were practice; the score was the mean correct-trial RT for the remaining 50 trials. Trials for which RT was below 100 ms or above 1000 ms were not included.

Procedure. The ability tests were administered in the first four sessions of the experiment. The order of administration, presented in Table 3, was invariant across subjects. Each session lasted approximately 90 minutes, allowing for short breaks between tests.

Semantic Category Search Task - Training

Stimuli. Stimulus-set items were the semantic categories Four-footed Animals, Weather Phenomena, Fruits, Types of Cloth, Kinds of Weapons, Kinds of Money, and Articles of Furniture (Battig and Montague, 1969). Six high associates were chosen from each category; the rankings ranged from one to seven. According to the norms collected by Collen, Wickens, and Daniele (1975), these categories were sorted together less than ten percent of the time and may thus be considered unrelated. The assignment of categories to conditions was counterbalanced across subjects by a Latin square.

Table 3. Experimental Procedure

Session 1 Monday	Information Sheet, Health and Medication Questionnaire
	Informed Consent and Overview
	Extended Range Vocabulary
	Computation Span
	Digit Symbol Substitution
	Simple Reaction Time
	Semantic Matching
Session 2 Tuesday	Eye Examination
	Miller Analogies Test
	Identical Pictures
	Mathematics Aptitude
	Making X's
	Listening Span
Session 3 Wednesday	Raven's Progressive Matrices
	Crossing Lines
	Number Comparison
	Information
	Semantic Access
Session 4 Thursday	Finding A's
	Controlled Associations
	Alphabet Span
	Letter Sets
	Synonym Matching
Sessions 5 - 9 (Fri-Thurs)	1200 Trials Category Search Task (600 CM and 600 VM)
Session 10 Friday	300 Trials CM Category Search Task
	420 Trials CM Reversal
	420 Trials New CM
	Experimental Debriefing

Visual angle for words and displays was calculated based on an average viewing distance of 46 cm from the screen. The visual angle subtended by the longest word was $.59^{\circ}$ in height and 1.58° in length. The visual angle from the center of the screen (the location of the focus cross) to the center of any word was 1.58°. The entire display (four words) subtended 1.98° in height and 4.75° in length.

Design. The within-subject independent variables were: (a) Display Size: two, three, or four words; and (b) Training Conditions: CM or VM. The primary dependent variable was RT. Trial-level accuracy was also recorded. Subjects were instructed to maintain a 95-percent accuracy rate.

Procedure. An experimental trial consisted of the following sequence of events. Subjects were presented with the memory set of one category label and allowed to study it for a maximum of 20 seconds. Once the memory set had been encoded, subjects pressed the space bar to initiate the trial. A plus sign then appeared in the center of the screen to allow the subjects to localize their gaze. After 500 ms, the display set was presented, this set consisted of up to four words presented in two rows of two words. (Figure 3 is a representation of the trial sequence.) Subjects were to indicate the location of the target item by pressing the key corresponding to the location: upper left (UL), upper right (UR), lower left (LL), or lower right (LR). There was a one-to-one correspondence between the location of a word on the screen and the location of its response key (labelled UL, UR, LL, LR) on the number pad. A target was present for every trial. Two, three, or four words were presented in each display. For those trials of less than four words, a placeholder was used which consisted of five characters (#@$&%). The placeholders were used to ensure that display load effects were due to semantic load

Memory Set	Focus Cross	Display Set	Response	Feedback
FLOWER	+	CHAIR TULIP + TABLE SOFA	UL $\boxed{\text{UR}}$ LL LR	Correct, Your RT was ...
WEAPON	+	RIFLE PEAR + APPLE #@$&%	$\boxed{\text{UL}}$ UR LL LR	Correct, Your RT was ...
ANIMAL	+	#@$&% WOLF + NYLON #@$&%	UL UR $\boxed{\text{LL}}$ LR	Error, the correct response was ...
subject initiation (or 20 s)	500 ms	subject response (or 6 s)		800 ms

Figure 3. Examples of Trial Sequences (Display Sizes 2, 3, and 4)

and not lateral masking (Fisher, Duffy, Young, and Pollatsek, 1988).

Subjects received the following performance feedback. After each correct trial, RT was displayed in milliseconds. After each incorrect trial, an error tone sounded and the correct word was displayed. Following each block of trials, subjects received their mean RT and accuracy for that block. If accuracy fell below 92 percent in any block, a message was displayed encouraging the subject to respond more carefully. If accuracy was 98 percent or higher, a message was displayed encouraging the subject to respond faster. Subjects were instructed to maintain a 95-percent accuracy rate while responding as quickly as possible. Before each daily session, the experimenter privately discussed with each subject his/her performance in the preceding days.

A 95-percent accuracy rate was chosen because it allowed subjects to make some errors. Thus, they were not trying to be "perfect," rather, they were aiming for speed while maintaining this accuracy level. Salthouse (1985) suggested that measuring RT differences at a constant accuracy level is a viable method for controlling speed/accuracy tradeoffs. Additionally, for accuracy rates above 90 percent, set size functions are unaffected, even when subjects were instructed to trade speed for accuracy (Shiffrin, 1988; Sternberg, 1975).

Subjects were trained on the semantic category search task in five 90-minute sessions (see Table 3). They first received written instructions on the task which included a list of the categories and words to be used. Subjects received practice in two conditions: (a) Consistent Mapping - target and distractor items were drawn from distinct stimulus sets, and (b) Varied Mapping - target and distractor items were drawn from the same stimulus set with replacement across trials. Each subject was assigned one

category as the CM target set and another category as the CM distractor set. The remaining five categories served interchangeably as targets and distractors in the VM condition. Each session consisted of 20 blocks of 60 trials per block (20 trials per display size). There were ten blocks each of CM and VM (all sessions began with a CM block; VM and CM blocks alternated thereafter). After each block of 60 trials, subjects were given the opportunity to take a short break (self-paced) to rest their eyes. Each subject completed a total of 3000 CM trials (1000 at each display size) and 3000 VM trials (1000 at each display size).

Semantic Category Search Task - Transfer

The final session of the experiment consisted of a transfer condition designed to assess the degree of target strengthening and distractor learning in the CM practice condition. Subjects first received five blocks of CM practice (300 trials--100 trials at each display size). These practice trials used the same CM target/distractor pairings as those used during the practice phase of the experiment. Following these trials, the subjects entered the transfer phase. There were two conditions in this phase: (a) CM Reversal - CM target and distractor roles were reversed (i.e., the previous CM targets became the distractors and the previous CM distractors became the targets) and (b) New CM condition - created by pairing two of the former VM categories in a consistent mapping. Seven 60-trial blocks of CM Reversal and seven 60-trial blocks of New CM were presented alternately for a total of 420 CM Reversal trials (140 at each display size) and 420 New CM trials (140 at each display size). The procedure for individual trials was the same as during practice.

Design. The within-subject independent variables for the transfer phase were: (a) Display size: two, three, or

four words, and (b) Training/Transfer Conditions: CM, CM Reversal, or New CM. The primary dependent variable was RT; subjects were instructed to maintain accuracy at 95 percent.

Statistical Procedures

The first step in the data analysis process was a normative assessment of the ability data and the improvements on the search task across practice. RT, accuracy, comparison slope estimates (increase in RT for corresponding increase in display size), and intercepts were all analyzed as a function of Training Condition (CM, VM) x Display Size (2, 3, 4) x Trial Block (1-50) using a repeated measures analysis of variance (ANOVA). RT, accuracy, comparison slope estimates, and intercepts were also analyzed for the transfer sessions. Performance in the CM Reversal and New CM conditions was contrasted with final-level CM performance. Effect sizes in the form of omega squared (ω^2) were calculated based on the formula provided by Dodd and Schultz (1973) for repeated measures designs.[5] Descriptive statistics (i.e., means and standard deviations) were tabulated for all the ability tests. Part 1/Part 2 reliability coefficients were obtained when possible.

The second phase of the analysis involved causal modelling of latent variables. The latent variables represent unobserved constructs which are not directly measured but are assessed through the measurement of manifest variables presumed to be indicators of the construct (Bollen, 1989; Everitt, 1984). In the present

[5]Although ω^2 is an index of the relative magnitude of an effect (relative to the total reliable variance), the absolute values of the effect sizes must be interpreted with caution because even very small effects can represent theoretically interesting manipulations (Keppel, 1973). Furthermore, the use of a multidimensional qualitative independent variable can influence the value of ω^2 (Kirk, 1982) and age is certainly a multidimensional qualitative variable (Schaie and Hertzog, 1985).

experiment, the latent constructs are the ability and performance factors. Multiple indicators of the latent constructs were used because each individual indicator was assumed to be measured with error. A latent factor can be extracted from multiple indicators that is essentially attenuated for measurement error (Joreskog and Sorbom, 1986).

Prior to the analysis of the relationships among the latent variables (i.e., the structural model), it was necessary to test the relationships between the manifest variables and the latent constructs (i.e., the measurement model). Analysis of the measurement model is basically a confirmatory factor analysis. Only after the goodness-of-fit of the measurement model has been demonstrated can structural models can be assessed (James, Mulaik, and Brett, 1982).

Best-fitting measurement models were determined separately for the ability data and search data using the maximum likelihood estimation procedure in LISREL VI (Joreskog and Sorbom, 1986). This procedure provides a chi-square (χ^2) statistic which allows assessment of the overall fit of the model. The LISREL goodness-of-fit index (GFI) and Bentler's (1990) comparative fit index (CFI) will also be presented to provide quantitative indices of the fit of the overall models. GFI is the result of comparing the hypothesized variance/covariance matrix and the sample covariance/variance matrix. It is equal to one minus the proportion of the total variance that is due to error variance. This index is biased in small samples (James et al., 1982). CFI represents the change in model fit between two nested models. In the present analysis, comparisons will be made between a "null" model which represents the most restricted model (the variables are assumed to be mutually uncorrelated) and the less restricted models which

are proposed to fit the data (e.g., a six-factor model of abilities). According to Bentler (1990), CFI is a relatively unbiased index even in small samples.

LISREL also provides modification indices and normalized residuals which allow assessment of the fit of individual parameters in the model. Modification indices provide an indication of the amount of change in the fit function that will occur if a particular parameter is freed. Normalized residuals are useful for model fitting because those larger than 2.0 indicate specification errors in the model (Hill, 1987).

Covariance matrices were used in all the models rather than correlation matrices. Because correlations are obtained by scaling variables according to sample variances, they are not directly comparable across samples. Furthermore, the likelihood ratio statistic is only an estimate of χ^2 when covariance matrices are analyzed, not correlation matrices (Alwin, 1988). The use of covariance matrices requires that a scale be determined for the latent variables. This was accomplished by arbitrarily fixing one factor loading of each latent variable to 1.0 (Joreskog and Sorbom, 1986).

The initial measurement model of abilities was based on the predicted factor structure shown in Figure 2. Adjustments to this model were made based on an assessment of the zero-order correlations in conjunction with the normalized residuals and modification indices of this initial model. A hierarchical model was then fit to the data and goodness-of-fit was assessed by comparing the base model with the hierarchical model. In the base model, the ability factors are freely intercorrelated whereas in the hierarchical model the intercorrelations are constrained to be a function of the higher-order factor (in this case, G). If there is not a significant χ^2 difference between the

models, it can be assumed that the hierarchical model appropriately represents the intercorrelations of the factors.

The first block of each session served as the basis for comparisons across sessions, thus avoiding the potential problem of within-session effects. There were 600 trials each of CM and VM per session, thus there was the potential that ability/performance relationships could change within a session. This is a particularly important concern for the first session during which practice effects are the greatest. Consequently, multiple blocks of trials from the first session were also analyzed separately to assess early changes in ability/performance relationships within that session.

For each set of search data, a null factors model was first run to serve as the baseline for CFI calculation (Bentler, 1990). Several hypotheses were tested to aid in determining the best-fitting model for the search data and reduce the number of parameters to be estimated. First, the hypothesis that the loadings of the display sizes could be constrained to be equal across sessions (or blocks in Session 1) without decreasing the fit of the model was tested. Constraining these parameters to be equal not only equated the relative influence of each display size on the performance factor, but also reduced the number of parameters to be estimated. χ^2 for the model in which the loadings were constrained to be equal across sessions is compared to χ^2 for the model in which the loadings are freely estimated. If there is not a significant difference in χ^2, the model in which the loadings are constrained equal may be retained.

The second model that was tested was the fit of a first-order autoregressive process to assess the simplex

structure of the data.[6] In other words, a model in which the covariances among the performance factors were freely estimated was compared to a model in which the covariances were constrained such that Session 1 predicted Session 2 which predicted Session 3 which predicted Session 4 which predicted Session 5 (Kenny, 1979). If the fit of these two models is not significantly different (as assessed by a χ^2 difference test), one can assume that a first-order autoregressive process appropriately represents the covariances among the performance factors. A model was also tested in which the coefficients of the autoregressive process were constrained to be equal across time. This model assesses whether the degree to which prior performance predicts future performance changes across practice (i.e., is the path coefficient from Session 1 to Session 2 equivalent to the path coefficient from Session 4 to Session 5). The best fitting measurement models for the search data are presented with longitudinally standardized factor loadings. Variances across sessions (or blocks in the Session 1 analyses) were pooled and used to obtain the scaling matrices (Hertzog and Schaie, 1986).

The measurement models for the ability data and search data were combined in structural equation models to determine the relationships between abilities and performance, both within the first session and across practice sessions. A series of models was assessed to investigate the influence of cognitive and speed abilities on visual search performance starting with a restricted pattern of regression and successively adding parameters. In addition to the null factors model, a null structural model is also presented. In this model, there are no paths

---

[6]Given that the variables are measured with error, this is actually a test of a quasi-simplex and not a perfect simplex (Joreskog, 1970).

from abilities to search but the hierarchical structure of abilities is fit along with the autoregressive process.

Initially, the influences on search performance of within-group age and sex were included as control variables. Subsequent models included the influences of specific latent factors (i.e., abilities) on performance across practice trials, beginning with the first 60 trials of Session 1. The order of inclusion of the abilities was hypothesis-driven primarily based upon Ackerman (1987), Fleishman (1972) and Kyllonen and Woltz (1989). The influence of G on early performance was assessed first, followed by PS, SMA, and WM. Additional influences of Gf and Gc were also assessed. Inspection of the modification indices, along with substantive hypotheses, guided the investigation of the abilities which influenced later performance (e.g., Sessions 2 through 5). Decisions to retain influential variables were made on the basis of the significance of the path coefficients as determined by the t-statistic provided by LISREL. Thus, the approach was exploratory in nature and the goal was the development of models of ability/performance relationships which may be tested in a confirmatory manner in subsequent experiments (Hertzog, 1985).

In addition to the direct effects, total effects are also provided for the final models chosen to represent the ability/performance relationships for young and old adults, under conditions of CM and VM practice, within the first session and across practice sessions. These values are standardized by multiplying the total effect in the LISREL output by the ratio of the standard deviations obtained from the computed (implied) covariance matrix of the endogenous latent variables. Total effects represent the combined direct and indirect (mediated) effects of exogenous, independent variables to endogenous dependent variables

(Kenny, 1979). Note that the total effect between two variables is not equivalent to the correlation between two variables because the effect estimate does not include spurious relationships or the influence of intercorrelations among exogenous variables (Alwin, 1988; Kenny, 1979).

RESULTS AND DISCUSSION

Given the complexity of the data analyses, the results are presented in several stages along with summaries and discussions. First, a normative assessment of the search task improvements is presented. RT, accuracy, comparison slope estimates, and intercepts are all analyzed as a function of practice and training condition (i.e., CM or VM). These variables are also assessed as a function of the transfer manipulation and comparisons are made between performance on the New CM condition relative to the CM Reversal condition.

Descriptive statistics, *along with reliability* estimates, are then presented for all the ability measures. The structure of abilities and search performance was assessed through analysis of separate measurement models. These results are followed by the analysis of the structural model of the interrelationships among the ability factors and performance on the search/detection task.

Normative Results: Category Search Task

Training: Reaction Time

Correct-trial RTs[7] are presented in Figure 4. Display Sizes 2, 3, and 4 are plotted separately for each block of practice in CM and VM.

[7]RTs below 100 ms and above 4000 ms were not included in the analyses.

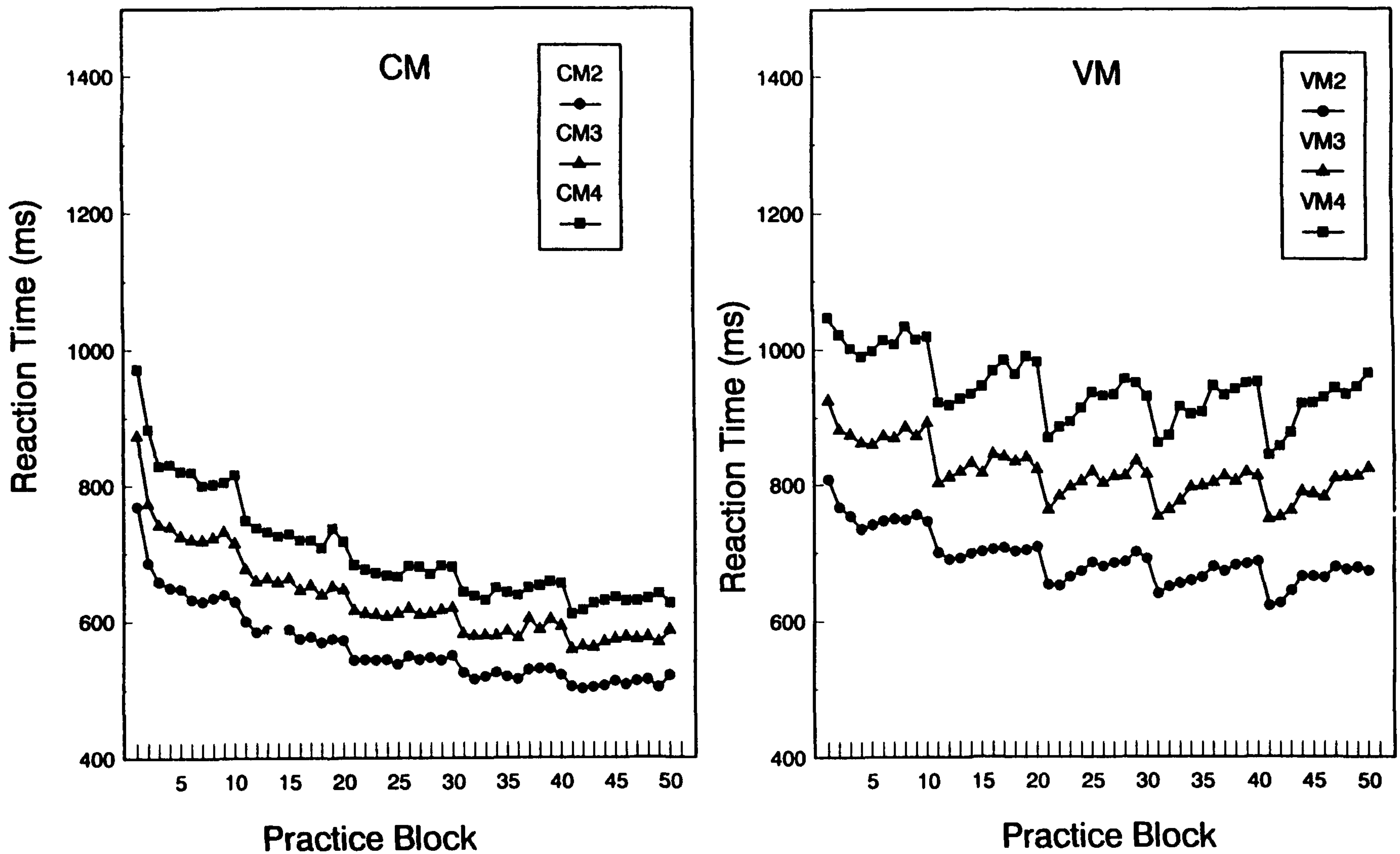


Figure 4. Reaction Times Across Practice (60 Trials Per Block)

An important consideration is the difference between CM and VM practice. It is important to determine if effects for small, restricted-subject samples generalize to large, relatively heterogeneous samples. There are differences in the amount of improvement for CM and VM as well as for the three display sizes. Analysis of these data revealed a significant Training Condition x Display Size x Practice interaction ($\underline{F}(98, 6762) = 2.89$, $\underline{p}<.0001$, $\chi^2 = .0008$). RT performance improved more overall in the CM condition relative to the VM condition, as revealed by the significant interaction of Training Condition x Practice ($\underline{F}(49, 3381) = 24.16$, $\underline{p}<.0001$, $\chi^2 = .0110$). However, within CM, RT improved more for larger display sizes whereas in VM there was more improvement for Display Size 2 relative to the larger display sizes. The dissimilarity between CM and VM is also evident in the correlations between the two conditions (Figure 5). The correlation between CM and VM was 0.70 for Block 1, but for the last block of practice the correlation between CM and VM was only 0.51; this is a significant difference ($\underline{t}(67) = 6.02$, $\underline{p}<.001$).

Training: Comparison Slopes

CM and VM comparison slope estimates are presented in Figure 6 across blocks of practice. As is clear from Figure 6, CM and VM slopes differ ($\underline{F}(1, 69) = 489.06$, $\underline{p}<.0001$, $\chi^2 = .2882$). Furthermore, CM slopes decreased as a function of practice whereas VM slopes actually increased slightly. These CM/VM differences are supported by a significant Training Condition x Practice interaction ($\underline{F}(49, 3381) = 4.56$, $\underline{p}<.0001$, $\chi^2 = .0134$).

Training: Intercepts

The intercept values for the CM and VM practice conditions are presented in Figure 7 across blocks of practice. The intercept values are reduced as a function of

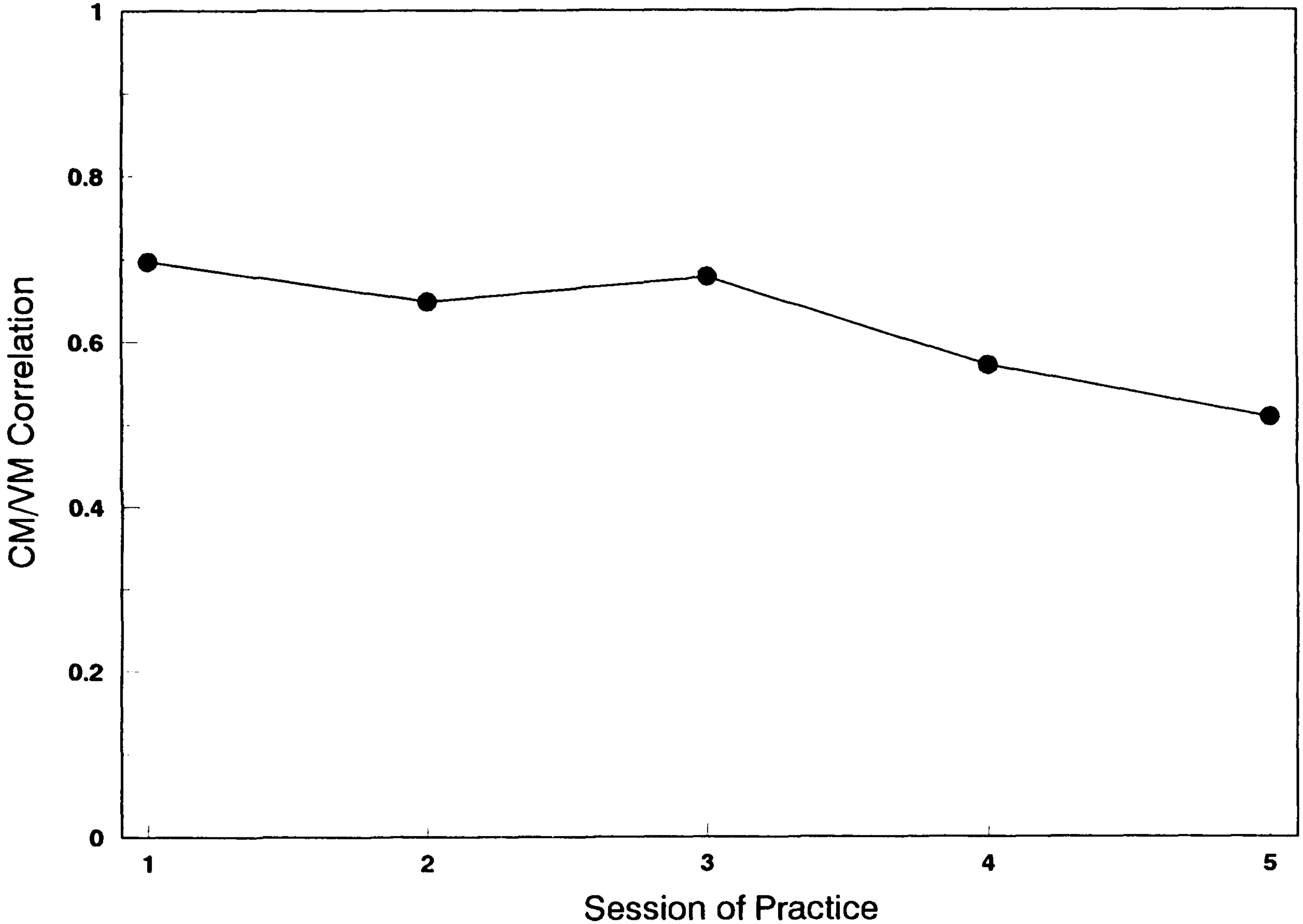


Figure 5. Correlation of Consistent Mapping with Varied Mapping Performance as a Function of Practice

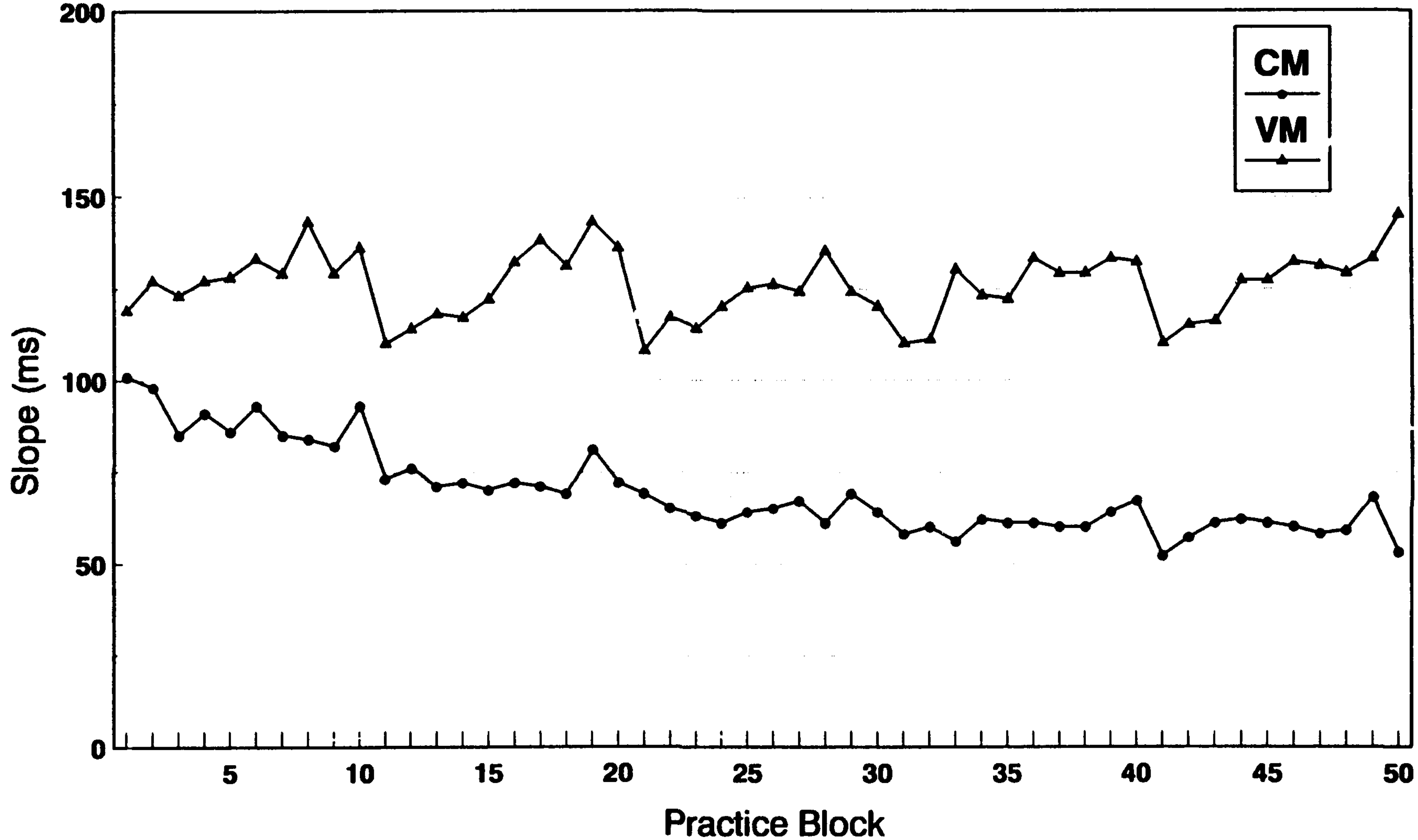


Figure 6. Comparison Slope Estimates as a Function of Practice Block
(60 Trials Per Block)

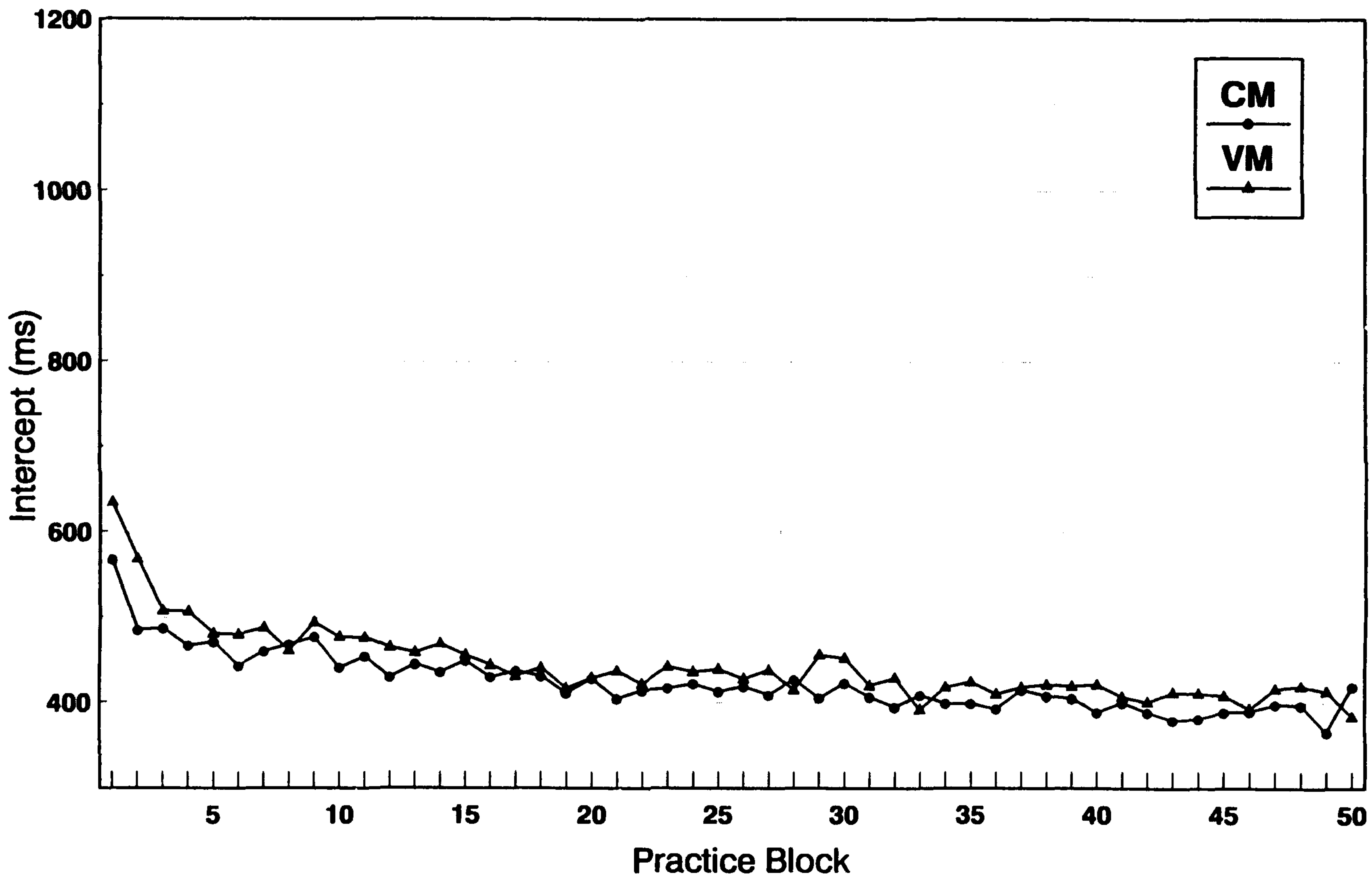


Figure 7. Intercepts as a Function of Practice Block (60 Trials Per Block)

practice ($\underline{F}(49, 6762) = 73.09$, $\underline{p}<.0001$, $\chi^2 = .0626$). Also, the intercept values differ only slightly between CM and VM (424 vs. 440).

Training: Accuracy

Mean accuracy rates are presented in Figure 8 for both CM and VM practice. As is clear from the figure, subjects were successful at maintaining the requested 95 percent accuracy rate. The overall accuracy rates for CM and VM were 95 and 94 percent, respectively.

Transfer: Reaction Time

The transfer session consisted of five blocks of CM followed by seven blocks each of CM Reversal and New CM (these blocks were presented alternately). The correct-trial RTs for the transfer session are presented in Figure 9. The introduction of the CM Reversal and New CM conditions *slowed* RT.[9] The significant Training/Transfer Condition x Display Size interaction ($\underline{F}(4, 552) = 127.48$, $\underline{p}<.0001$, $\chi^2 = .0098$) indicates that the amount of change in RT from practice to transfer is dependent on display size.

Transfer: Reaction Time Changes

The effects on RT for the transfer conditions were assessed by calculating difference scores separately for each display size: (a) CM Reversal Effect = CM Reversal RT - CM RT and (b) New CM Effect = New CM RT - CM RT. These formulas compare performance in the transfer condition to final-level CM performance. The difference scores for each block are presented in Figure 10 (CM Reversal) and Figure 11 (New CM). Subjects showed an increasing amount of disruption for increasing display sizes; however, the

[9] Only the first five blocks of New CM and CM Reversal are included in these analyses in order to equate the number of trials per condition.

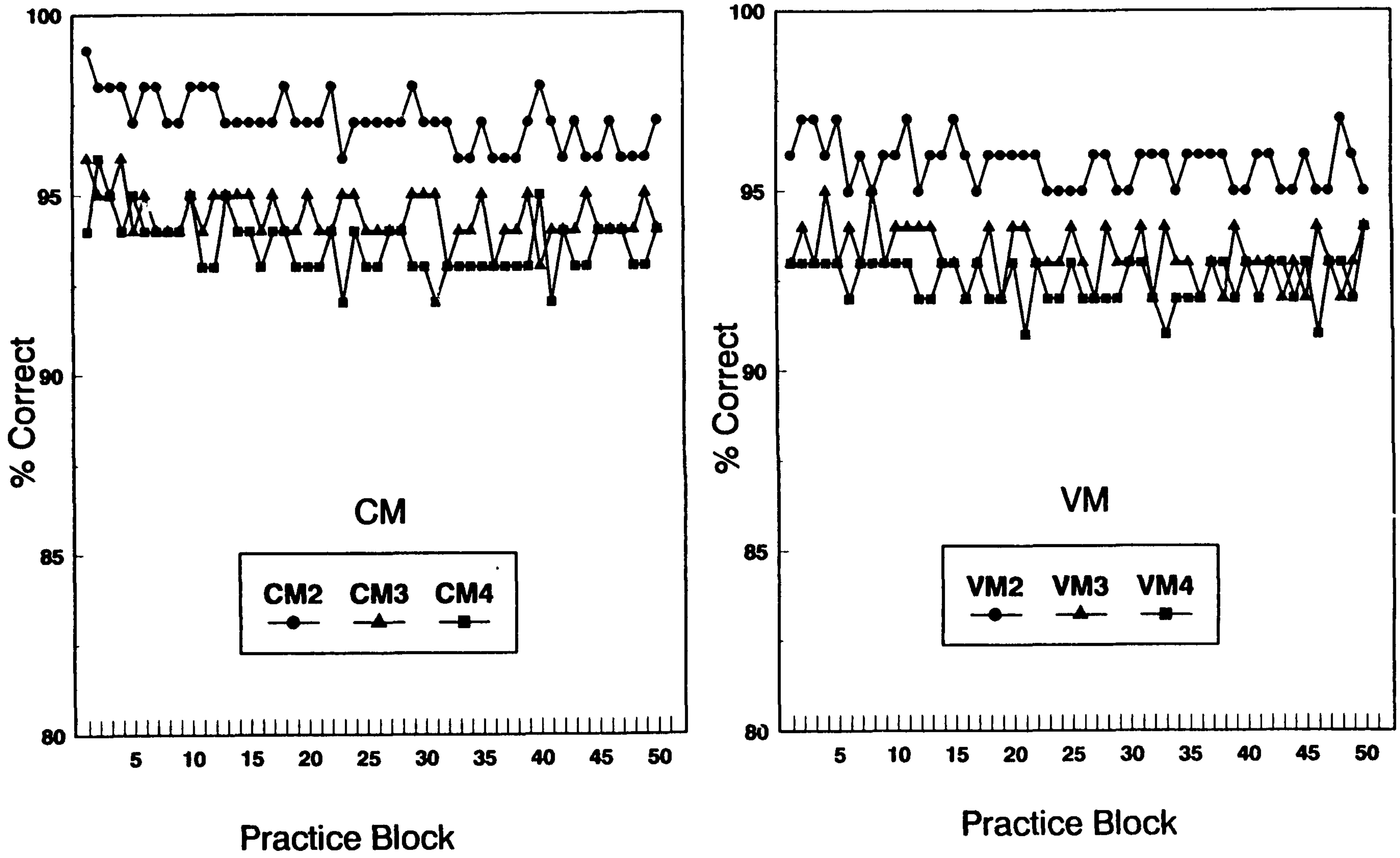


Figure 8. Accuracy as a Function of Practice Block (60 Trials Per Block)

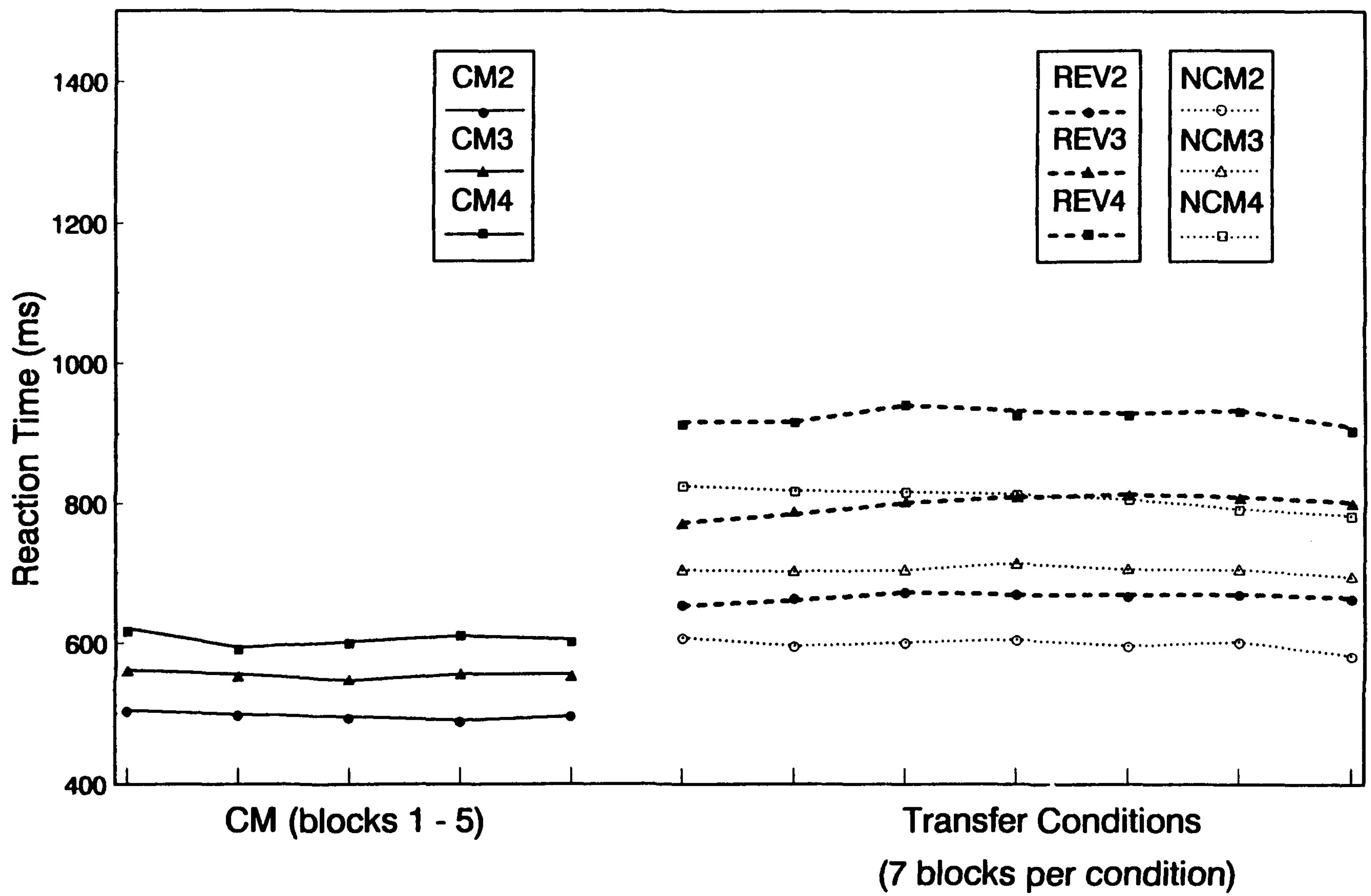


Figure 9. Reaction Times for Transfer Session

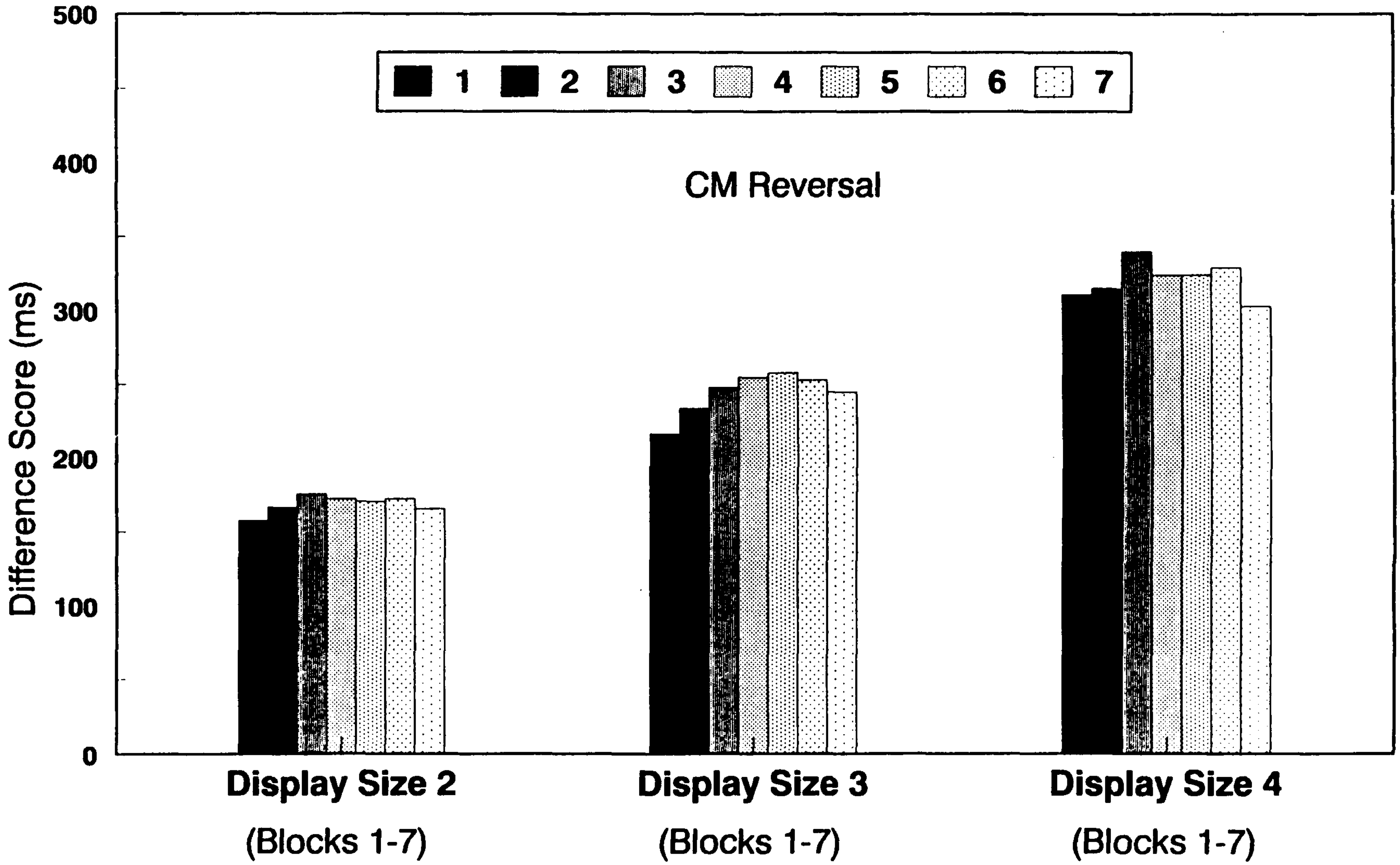


Figure 10. Absolute Reaction Time Differences (Consistently Mapped Reversal Reaction Times - Consistently Mapped Reaction Times)

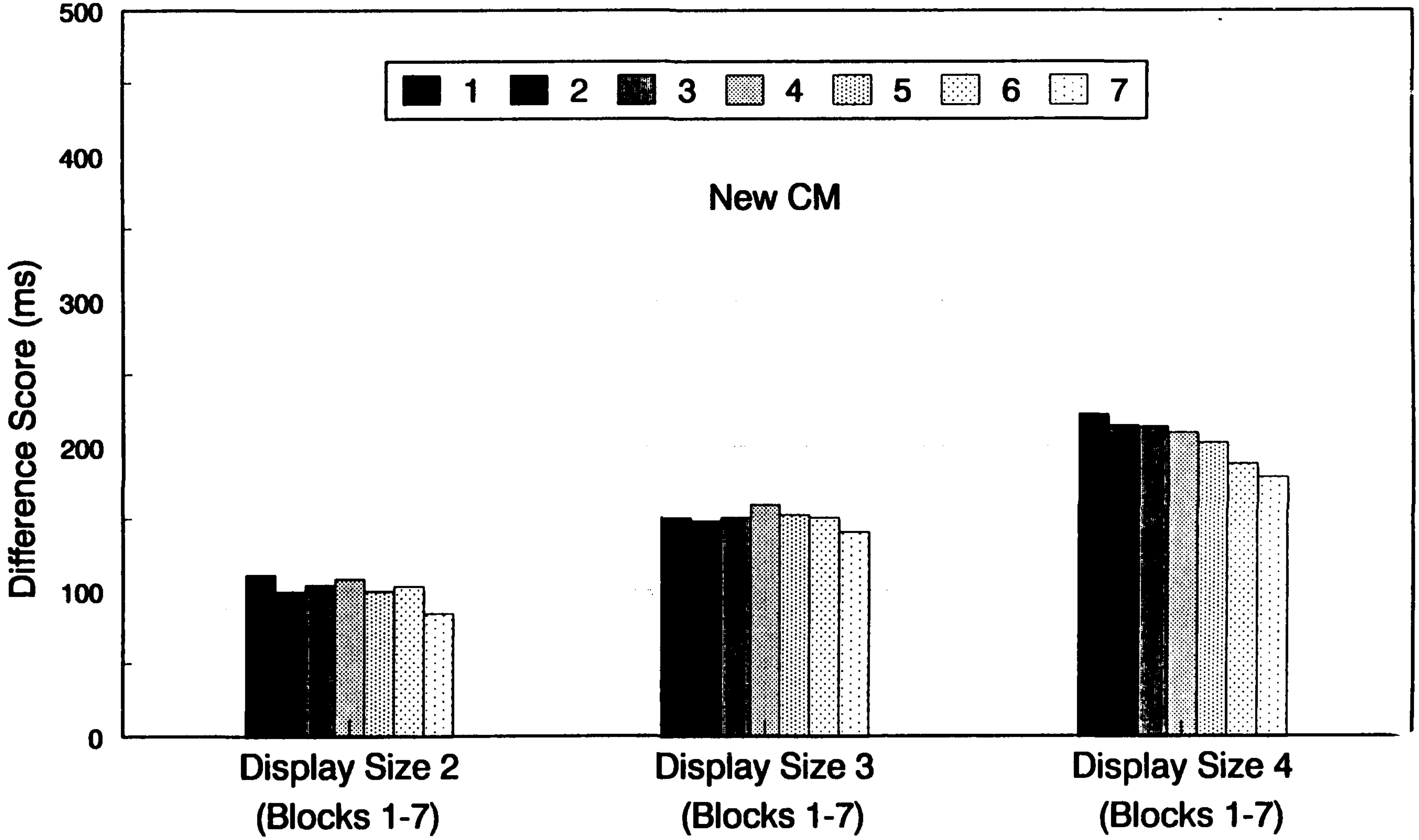


Figure 11. Absolute RT Differences (New Consistently Mapped Reaction Time - Consistently Mapped Reaction Time)

disruption is much greater for CM reversal than New CM. Reversal performance does not improve (i.e., subjects continue to be disrupted) throughout the transfer phase.

Transfer: Comparison Slopes

Comparison slope estimates for the last block of the trained CM condition may be compared with the slopes for the first block of each transfer condition in Figure 12. The CM slope for New CM was lower than that for CM Reversal; this is indicative of the larger disruption in the CM Reversal condition.

Transfer: Intercepts

Intercept values for the last block of the trained CM condition may be compared with the intercepts for the first block of each transfer condition in Figure 13. The subjects showed no significant change in intercept for the two transfer conditions relative to CM. Hence, we conclude that the disruption is due to changes required in the comparison processes.

Transfer: Accuracy

Mean accuracy rates are presented in Figure 14 for the transfer session. Subjects were very successful at maintaining the requested 95 percent accuracy rate, although there was a very slight decrease in accuracy for the CM Reversal condition. The overall accuracy rates were 95, 94, and 95 percent, respectively, for CM, CM Reversal, and New CM.

Discussion: Category Search Task

Performance Improvement

The CM and VM search data are consistent with previous small-sample studies of performance improvements in visual

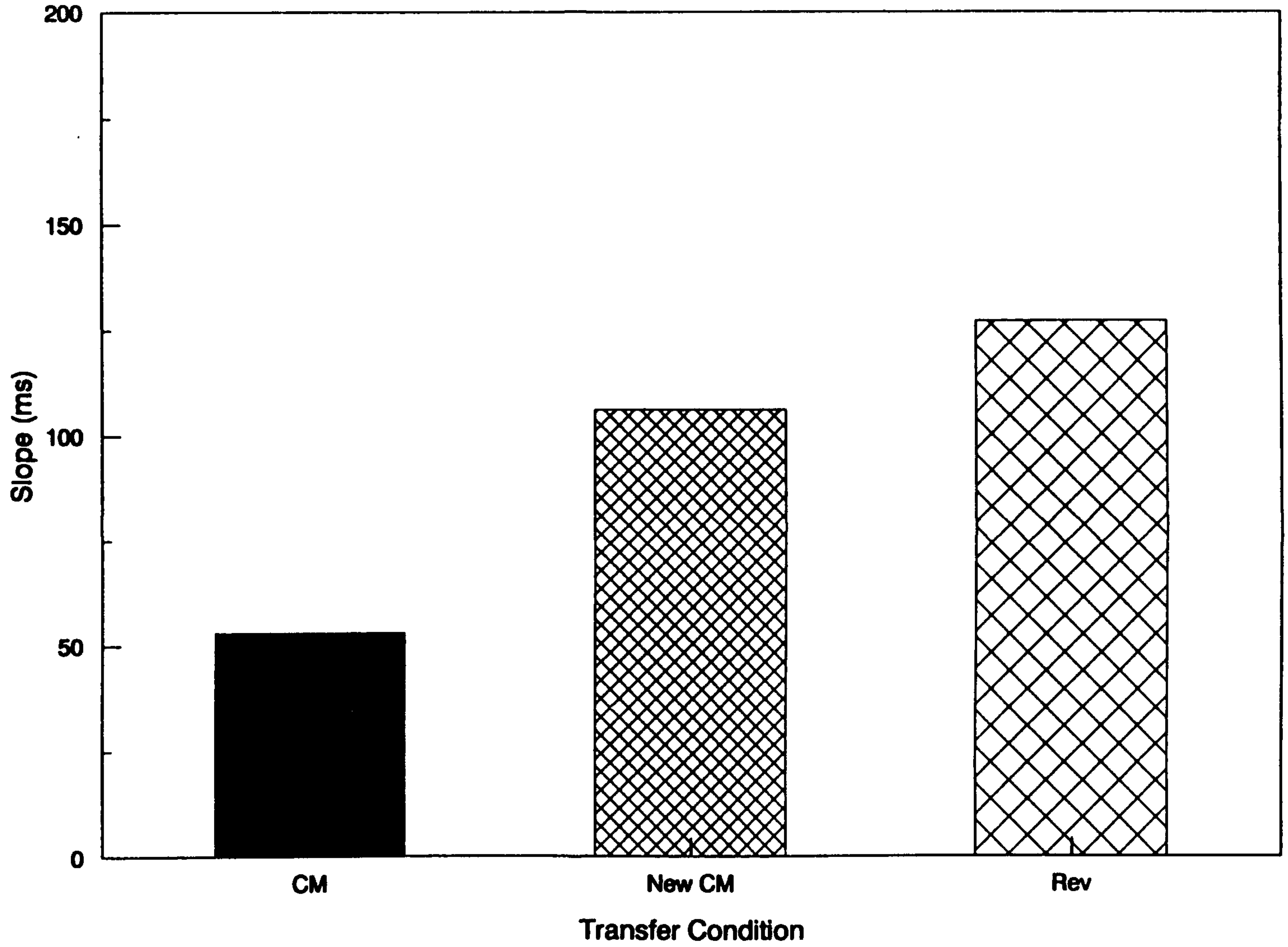


Figure 12. Comparison Slope Estimates for Transfer Conditions

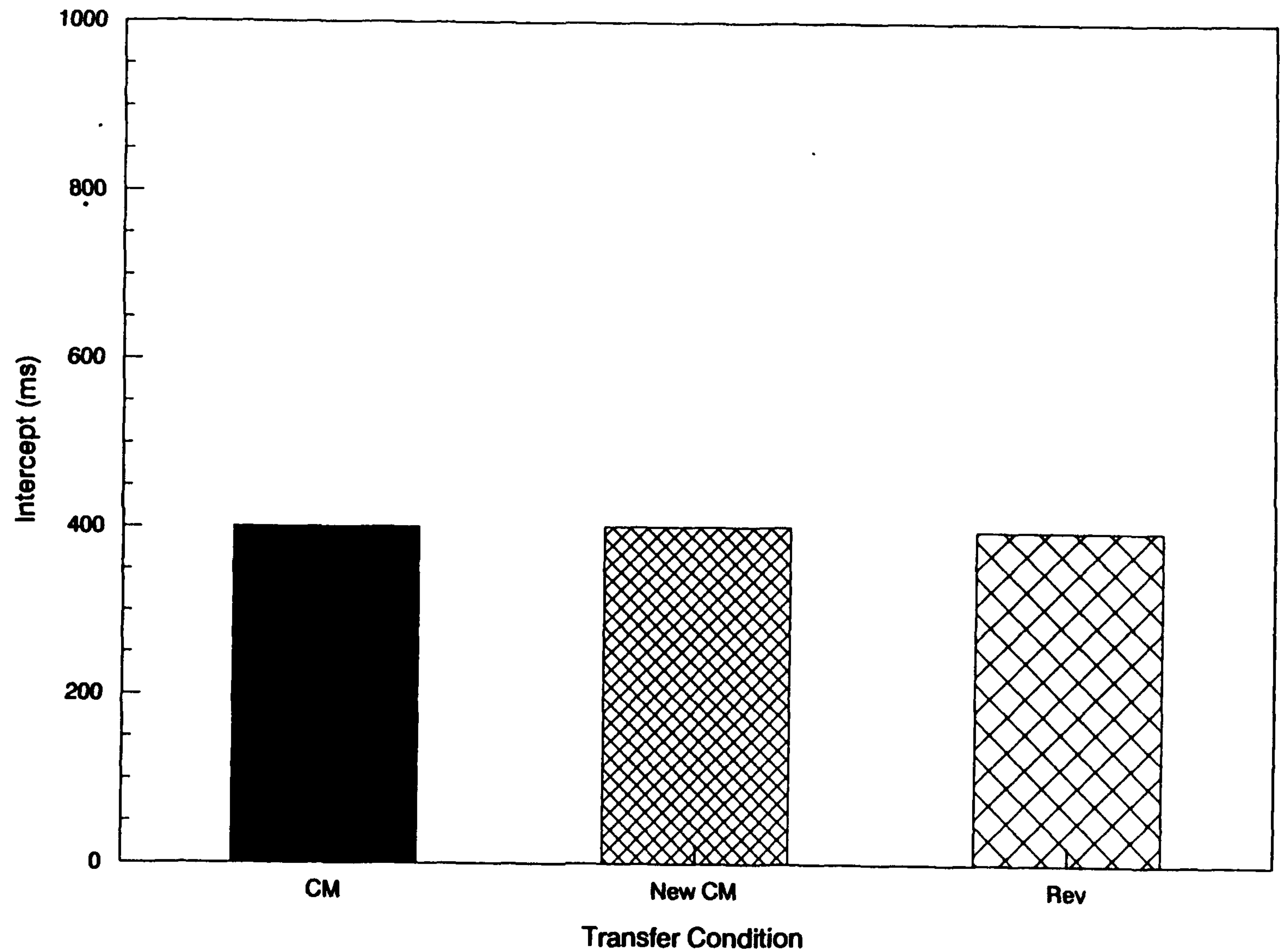


Figure 13. Intercepts for Transfer Conditions

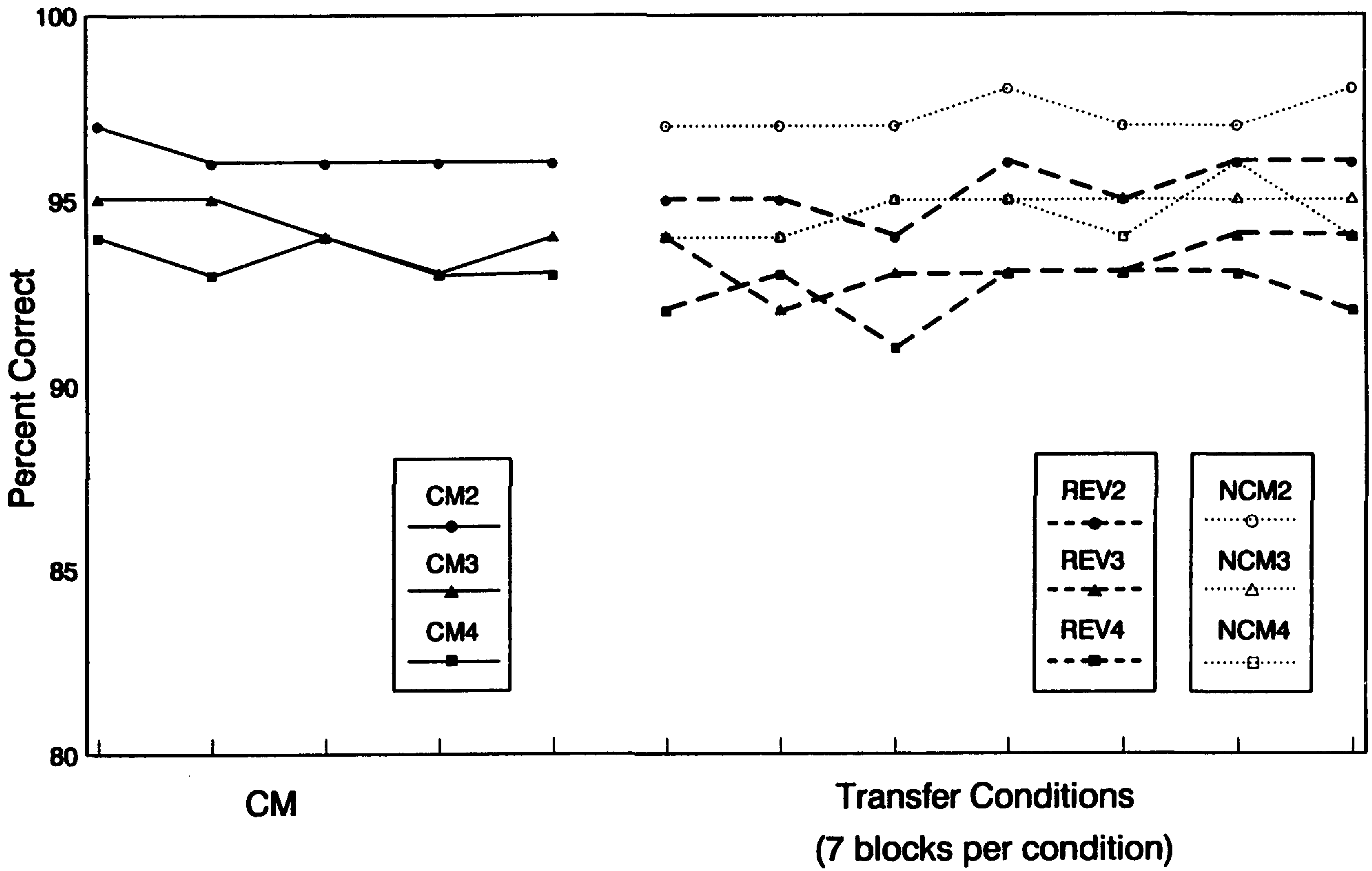


Figure 14. Accuracy Rates for Transfer Session

search. In the CM training condition, RT is reduced more with practice for higher display sizes. More improvement for the larger display sizes is consistent with the notion that an automatic response is developed as a function of CM practice. According to a strength-theoretic view of automatic response development in CM visual search, initial task performance requires a serial scan through the display for the target item (e.g., Shiffrin, 1988; Shiffrin and Schneider, 1977). If one must serially scan a display, scanning four items will require longer search times than scanning only two items. However, an increase in the attention-attraction strength of target items reduces the requirement for a serial search which, in turn, reduces RT for the larger display sizes.

The differential CM/VM performance improvement is consistent with the hypothesis that changes in performance as a function of practice may be driven by different mechanisms in CM versus VM. Subjects conformed to the improvement pattern predicted from previous category search results (e.g., Fisk and Schneider, 1983; Schneider and Fisk, 1984). That is, subjects improved more under CM practice in terms of both RT scores and comparison slopes. Comparison slopes were reduced more in CM because RTs were reduced for the larger display sizes with practice. Thus, subjects benefit from the consistency in the CM condition and, after 3000 practice trials, their performance is less influenced by increasing display sizes.

It is important to note that, although the subjects are less influenced by increasing display sizes, the comparison slope estimates are larger than zero in the CM search condition. This finding is consistent with previous visual search results (e.g., Czerwinski, 1988; Kristofferson, 1977; Shiffrin, 1991). There are several potential explanations for the 50-ms slopes observed for the CM condition in the

present experiment. For example, Shiffrin (1988) has proposed that slopes larger than zero represent a combination of serial and parallel search due to rechecking prior to the response. Another suggestion by Shiffrin (1988) is that the need for eye movements (e.g., to bring stimuli into foveal view) can impose a mechanical limit on search speed. Thus, in the present case, there might be a minimum time necessary to read the words because it is difficult to take in the entire display in a single fixation. Although there are only four items in the display, and Fisher (1982, 1984) has indicated that parallel processing is possible for displays of four elements, his model incorporated only single-character items. The words in the present experiment represent a more complex visual environment and also increase the visual angle subtended by the entire display. Thus, the 50-ms slope may be a function of the need to read each word serially to activate the category (i.e., 50 ms may be the minimum scanning time for these displays).

Within-Session Effects

The data presented in Figure 4 reveal a rather pronounced pattern of within-session increases in VM RT. While this pattern does not seem to be a typical result in experiments employing a visual search task, block-by-block data are not frequently reported. In fact, this pattern has been previously observed for VM performance (Ackerman, personal communication, May, 1991; Fisk and Schneider, 1981). This presently observed within-session effect cannot be explained on the basis of the current data. Anecdotally, however, subjects tended to work straight through the entire session without taking breaks. This performance style may be a contributing factor to the within-session effects. Another potential explanation is that subjects learn more and their performance speed increases in the CM condition.

It is relatively more difficult to do the serial search required in the VM condition. That is, there may be a form of proactive interference operating between the CM and VM trial blocks. These possibilities cannot be determined from the current data set but they do warrant further consideration. However, for current purposes, the within-session effects will be controlled in the structural models by only assessing ability/performance relationships for the first block of each session.[10] Session 1 is examined in more detail later, but the within-session effect is less striking in this session.

Summary: Transfer Performance

In the New CM condition, two of the former VM categories were paired such that one set served as the consistent target set and the other served as the consistent distractor set. Performance in the New CM condition was slowed relative to their previous CM performance. Recall that some disruption was predicted for new categories if CM practice resulted in the development of an optimal search strategy that was category-specific. However, any learning beyond such optimal search (i.e., automatic response development) can only be assessed in the CM Reversal condition.

Reversal of the previous CM targets and distractors resulted in a very large disruption (nearly 60 percent in some cases) and the amount of disruption is directly related to the number of "reversed" items in the display. Disruption at reversal is a well-replicated finding (Dumais, 1979; Fisk, Lee, and Rogers, in press; Rogers, 1989; Schneider and Shiffrin, 1977) and is presumed to indicate

[10]An overall analysis of variance was conducted on this subset of the data. The results pattern was essentially the same.

the amount of target/distractor differentiation that accrued during CM practice.

Summary and Implications of Search Results

Visual search performance for both the CM and VM training procedures improves as a function of practice. Comparison slope estimates show more flattening in CM than VM. Transfer performance reveals that both the CM reversal and the New CM conditions were disrupted relative to trained CM performance. The crucial finding was more disruption in the CM Reversal condition than in the New CM condition. The reduced disruption in the New CM condition suggests that subjects learn general search strategies (e.g., scanning strategies, learning the location of the response keys, improved motor coordination, etc.) along with the development of an optimal search strategy which may be stimulus-specific but does not involve automatic response development. However, the larger disruption in the CM Reversal condition suggests that subjects also develop an automatic response and, when this automatic response is counterproductive (in CM Reversal), their performance is disrupted.

Normative Results: Ability Measures

Means and standard deviations for the ability tests are presented in Table 4; available reliability estimates are presented in Table 5. The actual scores on the ability tests are comparable to extant data (e.g., Botwinick and Storandt, 1974; Salthouse and Mitchell, 1990; and Schaie, 1984).

Table 4. Ability Measures - Means, Standard Deviations

Ability Measure	Mean	SD
Vocabulary	24.08	7.2
Analogies	28.17	4.8
Information	20.57	3 4
Associations	30.73	8.9
Math	15.20	5.1
Raven's Matrices	25.58	5.3
Letter Sets	23.01	3.4
Computation Span	52.38	18.6
Listening Span	49.91	17.8
Alphabet Span	44.70	12.3
Semantic Matching	968.94	170.6
Synonym Matching	793.43	97.6
Lexical Access	567.44	65.6
Identical Pictures	81.95	11.3
Number Comp.	28.84	5.5
Finding A's	64.50	14.4
Digit Symbol	73.28	9.0
Making X's	217.47	19.0
Crossing Lines	113.56	13.7
Simple RT	223.67	34.4

Table 5. Ability Measures - Reliability Estimates

Ability Measure	Reliability Estimate
Vocabulary	.87[a]
Identical Pictures	.90[a]
Math	.90[a]
Raven's Matrices	.91[b]
Number Comparison	.88[a]
Information	.91[c]
Finding A's	.90[a]
Associations	.69[a]
Letter Sets	.69[a]
Digit Symbol	.92[c]
Making X's	.94[a]
Crossing Lines	.88[a]
Simple RT	.88[a]
Semantic Matching	.78[a]
Synonym Matching	.75[a]

[a]Split-half adjusted with Spearman-Brown Prediction Formula
[b]From Raven's Manual (1977)
[c]From Matarazzo (1972)

Results: Performance/Ability Structure

Measurement Model: Ability Structure

The correlations between all the ability measures, age, and sex are presented in Table 6. Note that RT scores in Simple RT, Lexical Access, Semantic Matching, and Synonym Matching have been reflected such that a positive relationship between these tests and the paper and pencil tests always indicates superior performance in the same direction. An initial confirmatory analysis, which included all the ability tests, was conducted to assess the first-order factor structure presented in Figure 2, although with freely correlated factors (i.e., no higher-order G factor was specified). The fit of the model was: $\chi^2(155, N=70) = 232.38$, $p<.000$, GFI=.758. An inspection of the modification indices and residuals along with the zero-order correlations revealed that several of the ability tests were not behaving as predicted. Given that the goal of this stage of the analysis was to produce a well-defined measurement model of abilities and reduce the number of parameters if possible, several of the ability tests were trimmed for the following reasons. The Controlled Associations test, which was proposed to measure Gc, did not correlate with the other Gc measures (i.e., Information, Vocabulary, and Analogies). Presumably, the Controlled Associations test is a measure of "verbal fluency" which is separable from more general Gc.

The raw scores for the Identical Pictures test revealed a ceiling effect for some of the young adults. The Educational Testing Service time limit was used and some subjects were able to complete the test with time to spare. Consequently, their scores did not reflect the extent of their ability. Furthermore, the Identical Pictures test correlated higher with Raven's (a measure of Gf) than with the other measures of PS.

Table 6. Ability Correlation Matrix[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Raven's Matrices	1.0												
2. Letter Sets	.570	1.0											
3. Math	.446	.381	1.0										
4. Analogies	.388	.266	.506	1.0									
5. Vocabulary	.135	.202	.337	.519	1.0								
6. Associations	.132	.252	.309	.234	.459	1.0							
7. Information	.383	.186	.440	.657	.514	.026	1.0						
8. Listen. Span	.245	.223	.328	.346	.154	.184	.368	1.0					
9. Comput. Span	.335	.385	.302	.302	.130	.050	.265	.510	1.0				
10. Alpha. Span	.215	.173	.199	.285	.334	.275	.239	.431	.514	1.0			
11. Lexical[b]	.112	.246	.193	.278	.308	.118	.263	.135	.271	.310	1.0		
12. Semantic[b]	.102	.297	.087	.424	.387	-.016	.317	.147	.336	.387	.568	1.0	
13. Synonym[b]	.017	.130	.072	.245	.388	.141	.365	.142	.248	.270	.589	.572	1.0
14. Find A's	.112	.258	.033	.123	.200	.236	-.033	-.004	.083	.170	.188	.109	.129
15. Digit/Symbol	.098	.068	.024	.194	.198	.084	.131	-.088	-.041	.136	.248	.153	.266
16. Num. Comp.	.115	.270	.012	.121	.163	.303	.023	.031	.020	.199	.384	.200	.295
17. Ident. Pict.	.393	.304	.268	.128	.078	-.008	.166	.093	.099	.009	.214	.053	.023
18. Making X's	-.208	-.180	-.139	-.209	-.212	-.169	-.170	-.250	-.087	-.156	-.038	.202	-.034
19. Cross Lines	.069	.021	.013	.026	-.232	-.133	-.034	.062	.131	-.C77	-.078	.096	.120
20. Simple RT[b]	.078	.217	.199	.202	-.018	-.159	.168	.071	.328	.000	.315	.180	.152
21. Age	-.394	-.334	-.193	-.297	.207	.163	-.189	-.145	-.278	-.017	-.097	-.077	.108
22. Sex[c]	.269	.215	.433	.220	-.033	-.141	.223	.132	.252	-.208	.073	.042	.019
Mean	25.6	23.0	15.2	28.2	24.1	30.7	20.6	49.9	52.4	44.7	-567.4	-968.9	-793.4
SD	5.3	3.4	5.1	4.8	7.2	8.9	3.4	17.8	18.6	12.3	65.6	170.6	97.6

[a] Correlations > .23 are significant at .05 [b] RT scores were reflected. [c] Coded: Males-1, Females-0

Table 6. Ability Correlation Matrix[a] (cont.)

	14	15	16	17	18	19	20	21	22
14. Find A's	1.0								
15. Digit/Symbol	.445	1.0							
16. Num. Comp.	.411	.384	1.0						
17. Ident. Pict.	.267	.329	.306	1.0					
18. Making X's	.152	.184	.182	.282	1.0				
19. Cross. Lines	.132	.136	.087	.308	.456	1.0			
20. Simple RT[b]	-.060	-.240	-.211	.027	-.030	.142	1.0		
21. Age	.130	.092	.022	-.270	.035	-.247	.313	1.0	
22. Sex[c]	-.230	-.218	-.334	.239	.053	.210	.417	-.412	1.0
Mean	64.5	73.3	28.8	82.0	217.5	113.6	-223.6	20.8	.6
SD	14.4	9.0	5.5	11.3	19.0	14.7	34.4	4.0	.5

[a] Correlations > .23 are significant at .05

[b] RT scores were reflected

[c] Coded: Males-1, Females-0

The decision not to include Controlled Associations and Identical Pictures is not problematic for two reasons. First, selection of variables will only change the structure of the ability if that ability has subfactors. Also, deleting one test reduces the influence of a particular subability. Second, these two tests were not needed to define their respective factors. Even with their exclusion there remained three indicators per ability.

The three measures of PM did not coalesce to form a single factor. Making Xs and Crossing Lines formed one factor which might represent "movement speed" whereas Simple RT (SRT) formed a separate factor. SRT was most germane to the current focus but the zero-order correlations of SRT with the other ability measures revealed that there were no significant correlations nor did SRT correlate with search performance. As a result, this factor was not included in subsequent analyses for young adults.

Age and sex were included as control variables. There was an age range of 15 years; this range is sufficiently large that there could be significant age differences. Sex was included as a control variable because the zero-order correlations revealed that there were significant sex differences on some of the ability tests (e.g., math) and, by inference, there might be sex differences in the latent ability constructs.

A restricted factor analysis was conducted on the remaining ability measures with seven factors: Gc, Gf, WM, SMA, PS, age, and sex. The loadings of age and sex were fixed to 1.00 and their residual covariances were fixed to 0 because they were assumed to be measured without error. The metric of each remaining factor was defined by arbitrarily fixing one of the factor loadings to unity for each factor (Number Comparison for PS, Lexical Access for SMA, Information for Gc, Alphabet Span for WM, and Raven's for

Gf). The Mathematical Reasoning test was allowed to load on both Gc and Gf. This model fit the data adequately (χ^2 (95, N=70) = 92.74, p<.546, GFI=.877).

The final stage of defining the measurement model involved including a second-order G factor. The metric of G was defined by fixing the Gf loading to 1. An initial hierarchical model defined the G factor with five subfactors: Gf, Gc, WM, SMA, and PS (Figure 15, Panel A). Inspection of the resultant modification indices for this model revealed that fit could be improved through re-specification of the model. A second model was fit in which a separate semantic access (SA) factor was formed that loaded on Gc and PS (Figure 15, Panel B). This model was not appropriate because it yielded a negative variance for the SA factor. A third model was then fit in which SMA was defined as a subfactor determined by PS and Gc (Figure 15, Panel C). Analysis of the modification indices of this model did not reveal any changes which would greatly improve fit (i.e., there were no readily apparent specification errors). The overall fit of the model was not significantly worse than the base model (p=.10); this suggests that the interrelationships between the latent factors were modeled appropriately.

The final measurement model of abilities is presented in Figure 16. The factor loadings are standardized to allow comparisons across variables. The residuals represent one minus the squared multiple correlation provided by LISREL (i.e., the uniqueness). The fit of the model to the data was adequate (χ^2(106, N=70) = 112.17, p<.322, GFI=.858, CFI=.983).

Discussion: Ability Structure

With the exception of the tests that were excluded from the ability measurement model (Controlled Associations,

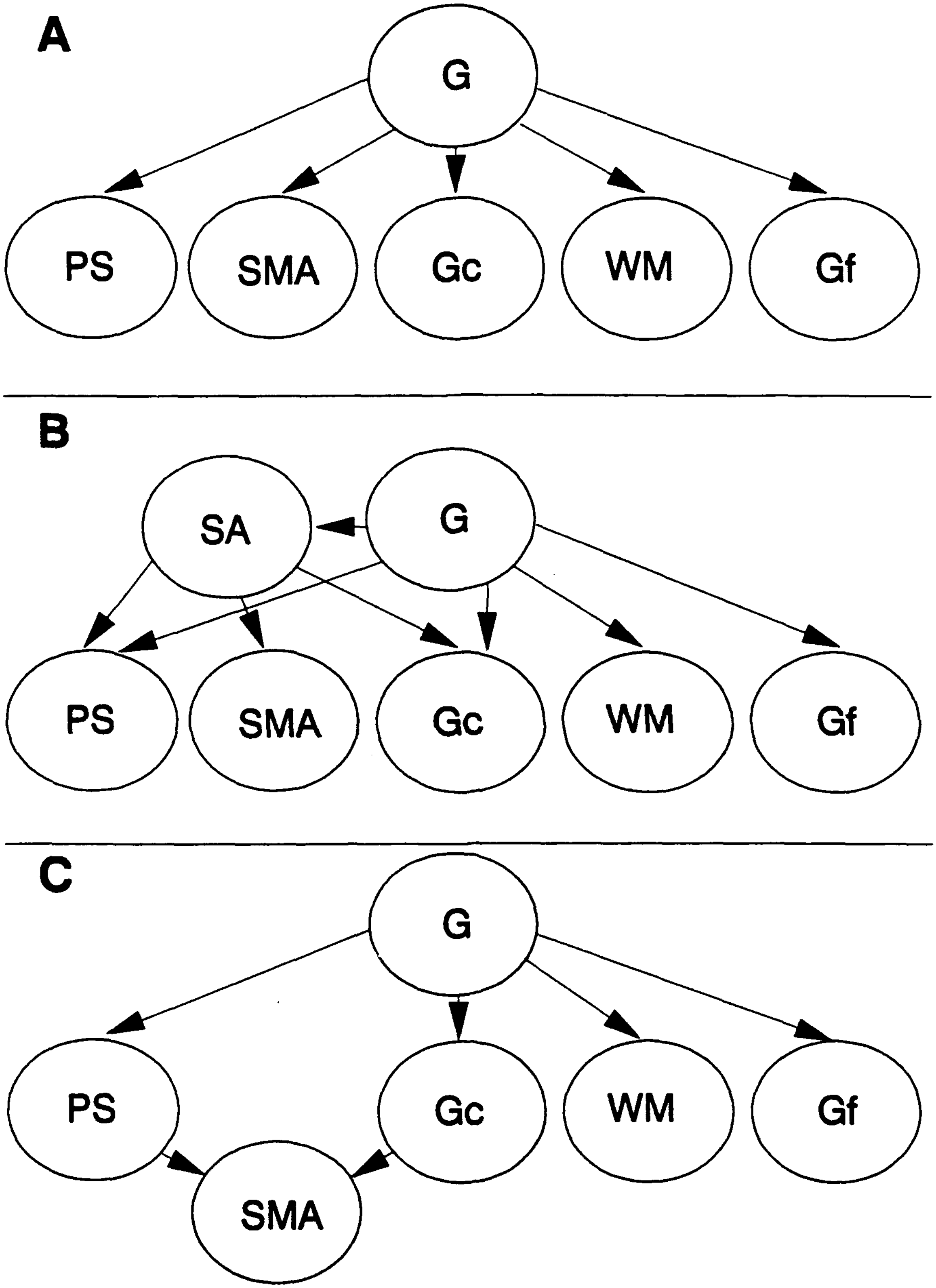


Figure 15. Sample Models for Ability Structure

Ability Structure

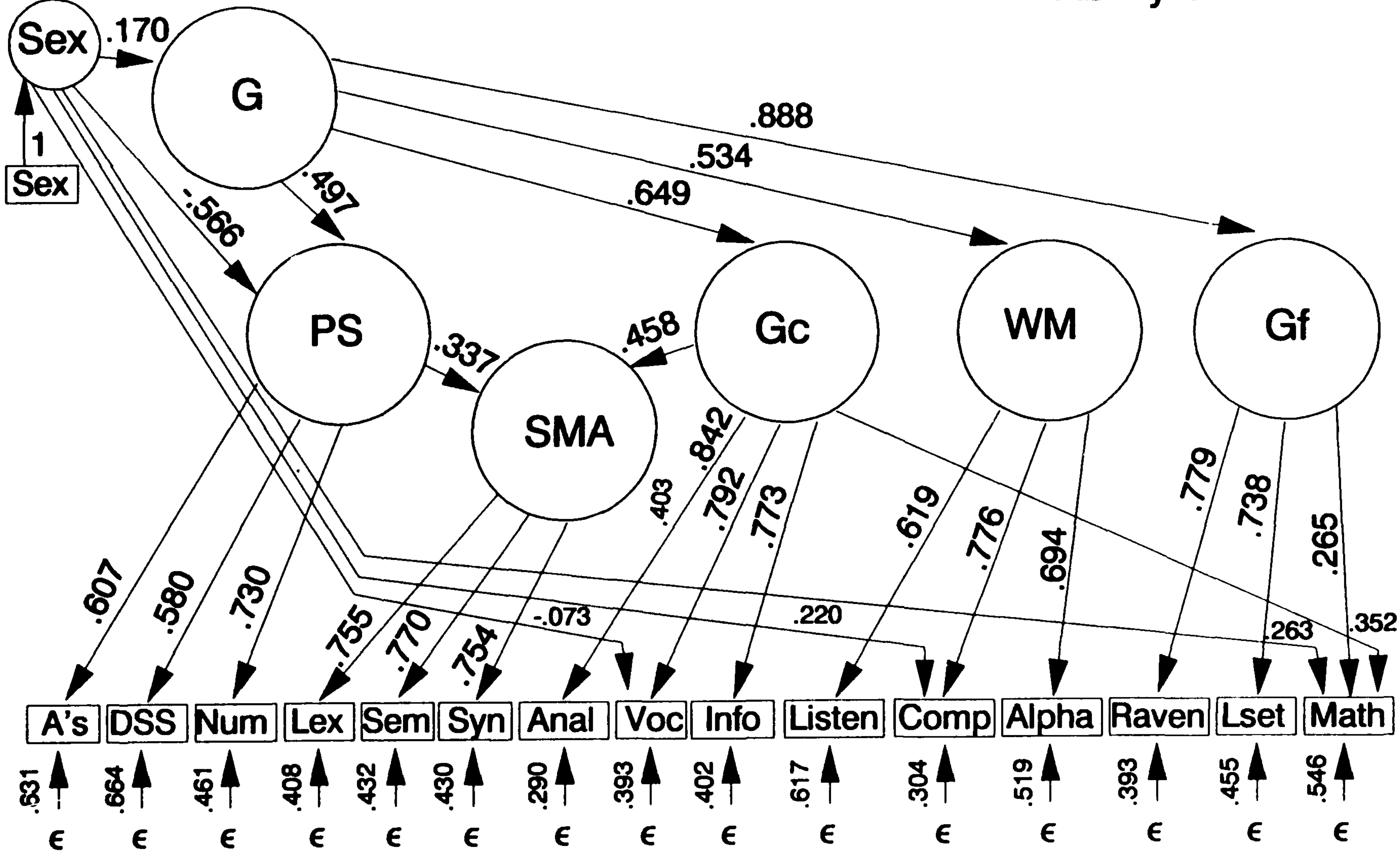


Figure 16. Measurement Model: Ability Structure

Identical Pictures, and the Psychomotor RT tests), the factor structure of the tests conformed to the predicted pattern. Table 6 (the intercorrelations of the measures in the battery) reveals a general pattern of convergent and discriminant validity such that the tests predicted to serve as indicators of a particular latent trait correlate highly with each other and are correlated to a lesser degree, if at all, with the indicators of other abilities. For example, the WM measures correlated higher with one another than with the other measures in the battery.

A restricted factor analysis fit the data well, although several modifications were needed to improve the fit. For example, the mathematical reasoning test was allowed to double-load on Gf and Gc. For this group of subjects, basic mathematical reasoning was less an indicator of fluid ability than an indicator of crystallized ability. However, almost 70 percent of the subjects were currently college students; that may have influenced this result. That is, if these individuals were using this particular reasoning ability frequently, the math test might have required less fluid, "on-line" reasoning.

There were also several demographic differences which, where included in the model, provided a better fit to the data. Males in the sample had an advantage in both the mathematical reasoning task and the computation span task (which required the manipulation of numbers). This finding may be an instantiation of the frequently observed male advantage in numerical facility (e.g., Maccoby and Jacklin, 1974; Minton and Schneider, 1980). There was also a slight female advantage on the vocabulary test. Finally, even within this group of individuals aged 17 to 31, there was an age-related influence on vocabulary which favored the older individuals.

In general, the ability structure is representative of findings in the literature. Gf has the highest loading on the higher-order G factor; this is consistent with Gustaffson's findings (1984). Gc and Gf were separable factors, as has been frequently reported by Cattell and Horn (Cattell, 1963; Horn 1982, 1985; Horn and Cattell, 1967). PS loads significantly on the higher-order G factor although to a lesser degree than Gf, Gc, and WM. Finally, SMA was best defined as a subfactor influenced by PS and Gc. This is consistent with the work of Hunt and his colleagues (Hunt, Frost, and Lunneborg, 1973; Hunt, Lunneborg, and Lewis, 1975).

Measurement Model: Consistent Mapping (All Sessions)

A CM performance factor for the first CM block of each session was defined using mean RT performance for Display Sizes 2, 3, and 4 as indicators. The metric of the factor was defined by fixing to 1.0 the loading of Display Size 2. The fit of this model, denoted CM1, is presented in Table 7 along with the null factors model for comparison. Model CM1 is clearly a major improvement over the null model. Post hoc evaluation of the fit of model residuals and inspection of the modification indices for θ_ϵ revealed that the fit might be improved if the errors for Display Size 2 were allowed to correlate across measurement occasions. Such autocorrelated residuals are common in longitudinal and repeated measures designs (Joreskog and Sorbom, 1977; Kessler and Greenberg, 1981). The fact that the residuals are correlated for Display Size 2 is concordant with the finding that subjects improved less for Display Size 2 (relative to Display Sizes 3 and 4). Autocorrelated residuals were included for Display Size 2 (see Model CM2 in Table 7) and yielded a significant improvement in fit (change in $\chi^2(10, N=70) = 23.73$, $p<.01$).

Table 7. Goodness-of-Fit Statistics: CM (All Sessions)

Model	χ^2	df	p	GFI	CFI
Null Model	1089.44	105	.000	.158	----
CM	1118.83	80	.003	.820	.960
CM2 (cov θ_ϵ D=2)[a]	95.10	70	.025	.858	.974
CM3 (cov θ_ϵ D=2, 3=, 4=)[b]	104.86	78	.023	.845	.973
CM4 (cov θ_ϵ D=2, 3=, 4=, simplex)[c]	112.24	84	.021	.834	.971
CM5 (cov θ_ϵ D=2, 3=, 4=, simplex, β =)[d]	134.65	87	.001	.810	.951

[a] Autocorrelated residuals for Display Size 2
[b] Factor loadings for Display Sizes 3 and 4 constrained to be equal over time
[c] Simplex pattern imposed on factor structure
[d] Beta coefficients constrained equal across time
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-fit Index

Model CM3 tested the hypothesis that the factor loadings for Display Sizes 3 and 4 could be constrained equal across occasions without loss of fit (Display Size 2 is functionally constrained equal across occasions because the factor loading was fixed to 1.0). A comparison of Model CM3 with Model CM2 supported the hypothesis that adding these constraints did not significantly change the fit to the data change in ($\chi^2(8, N=70) = 9.76$, $p<.30$).

Model CM4 represents an assessment of the simplex structure of the data by imposing a first-order autoregressive process. The absence of a significant change in fit relative to Model CM3 (change in $\chi^2(6, N=70) = 7.38$, $p>.30$) demonstrated that structuring the covariances as a first-order autoregressive process did not fit the data worse than allowing them all to vary. Model CM4 is presented in Figure 17 with the factor loadings standardized across longitudinal occasions. Observe that the values of the path coefficients between sessions increase over time. Model CM5 in Table 7 provides a test of whether the coefficients can be constrained equal without a significant loss of fit. A comparison of model CM5 with model CM4 reveals that constraining the simplex coefficients to be equal across time worsens the fit of the model change in ($\chi^2(3, N=70) = 22.41$, $p<.01$).

Measurement Model: CM (Session 1)

A CM performance factor was defined for each of the first and last two CM blocks of the first practice session. The results of this four-factor model are presented in Table 8 (Model CM1). This model is obviously an improvement over the null factors model. As in the analysis of all the sessions, allowing autocorrelated residuals for Display Size 2 (Model CM2) resulted in a significant improvement in the fit change in ($\chi^2(6, N=70) = 18.08$, $p<.01$). Model CM3 reveals that constraining the factor loadings of Display

CM Performance - Across Practice

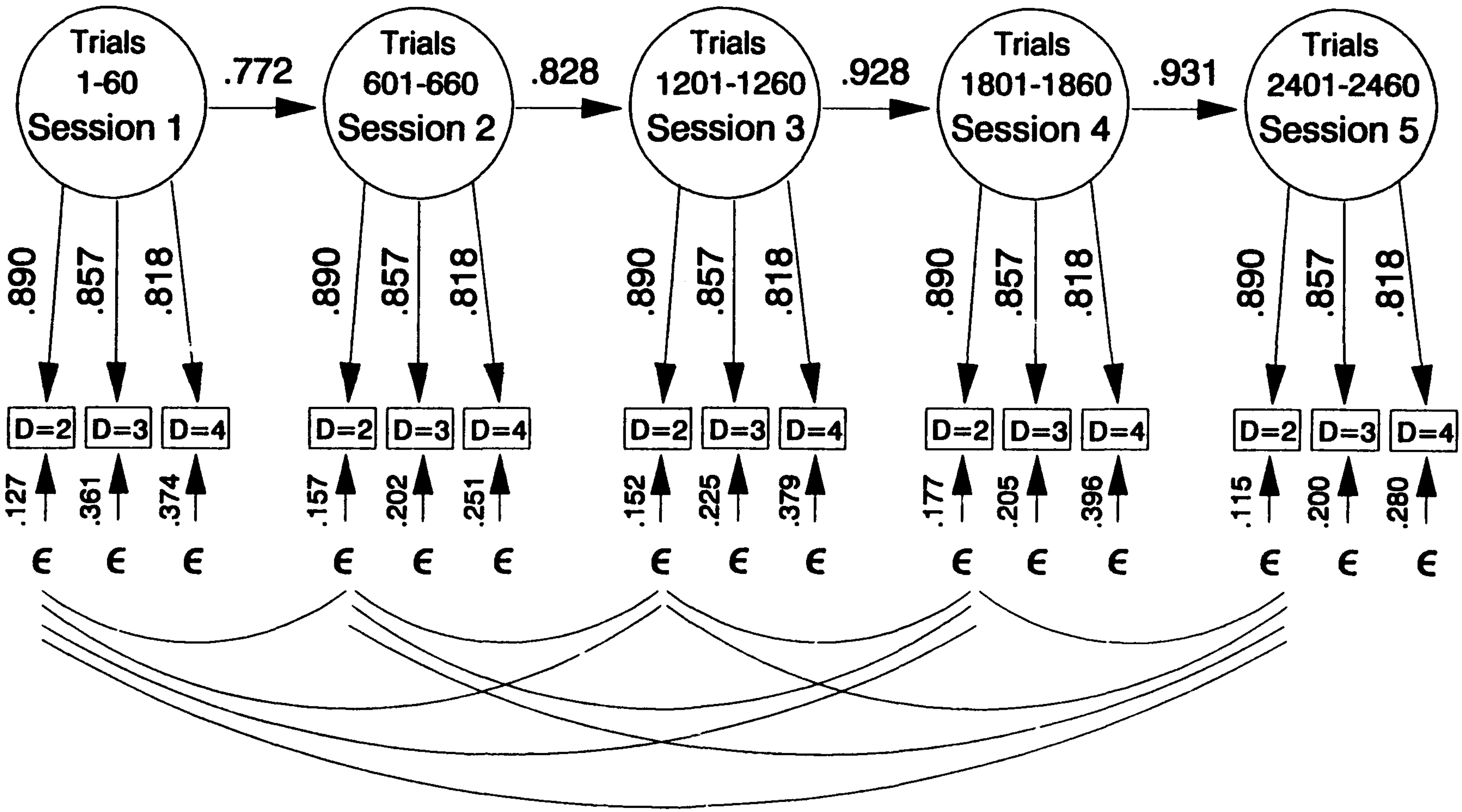


Figure 17. Measurement Model: Consistent Mapping (All Sessions)

Table 8. Goodness-of-Fit Statistics: CM (Session 1)

Model	χ^2	df	p	GFI	CFI
Null Model	833.77	66	.000	.193	----
CM	173.29	48	.011	.861	.967
CM2 (cov θ_ϵ D=2)[a]	55.21	42	.083	.894	.983
CM3 (cov θ_ϵ D=2, 3=, 4=)[b]	62.27	48	.081	.884	.981
CM4 (cov θ_ϵ D=2, 3=, 4=, simplex)[c]	72.37	51	.026	.863	.972
CM5 (cov θ_ϵ D=2, 3=, 4=, simplex, β =)	74.48	52	.022	.862	.971

[a] Autocorrelated residuals for Display Size 2
[b] Factor loadings for Display Sizes 3 and 4 constrained to be equal over time
[c] Simplex pattern imposed on factor structure
[d] Beta coefficients constrained equal across time
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-fit Index

Sizes 3 and 4 to be equal across blocks did not significantly worsen the fit of the model. A comparison of models CM3 and CM2 yielded a nonsignificant change in ($\chi^2(6, N=70) = 7.06, p<.30$). χ^2 for Model CM4 in Table 8 represents the fit of the first-order autoregressive process. Relative to Model CM3, there is a significant change in ($\chi^2(3, N=70) = 10.1, p<.02$). Thus, the first order autoregressive process does not represent the covariances as well as Model CM3, which allowed them to vary freely. However, the autoregressive process will be retained for the structural model to statistically control for prior-level performance when estimating ability/performance relationships across practice. Model CM4 is presented in Figure 18 with longitudinally standardized factor loadings.

Model CM5 tested whether the path coefficient from Block 1 to 2 could be constrained equal to the path from Block 9 to 10. (Due to the unequal interval, it was not appropriate to also constrain the path from Block 2 to 9 to be equal to the others; thus, this path was freely estimated.) A comparison of Model CM5 with Model CM4 yielded a nonsignificant change in $\chi^2(1, N=70) = 2.11$, $p<.15$.

Measurement Model: Variable Mapping (All Sessions)

A VM performance factor for the first VM block of each session was defined using mean RT performance for Display Sizes 2, 3, and 4 as indicators. The metric of the factor was defined by fixing to 1.0 the loading of Display Size 2. The fit of this Model, denoted VM1, is presented in Table 9. This model yielded a large improvement over the null factors model.

As in the CM condition, allowing correlated errors for Display Size 2 significantly improved the fit of the model,

CM Performance - Session 1

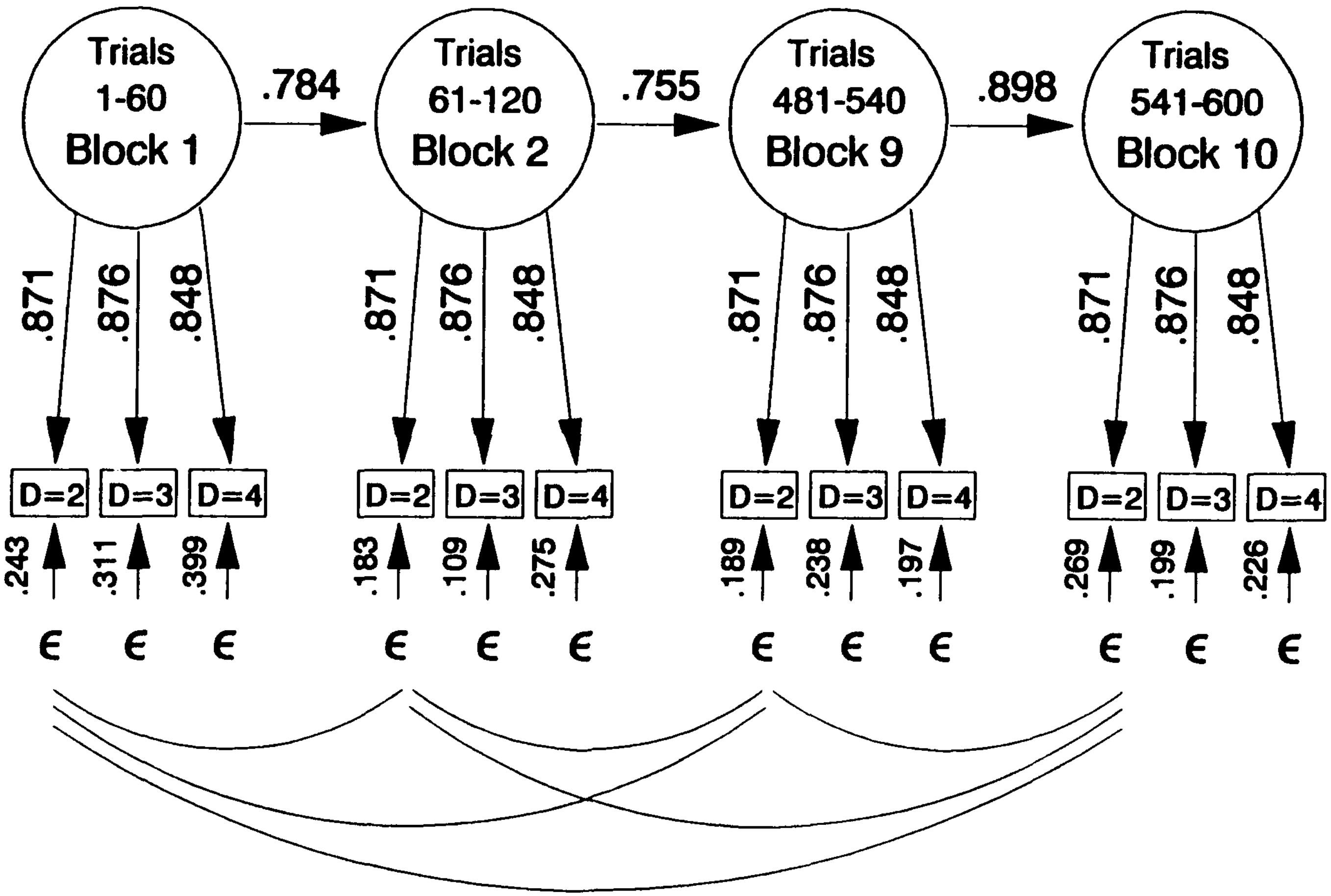


Figure 18. Measurement Model: Consistent Mapping (Session 1)

Table 9. Goodness-of-Fit Statistics: VM (All Sessions)

Model	χ^2	df	p	GFI	CFI
Null Model	1028.84	105	.000	.154	----
VM1	132.17	80	.000	.801	.944
VM2 (cov θ_ϵ D=2)[a]	96.97	70	.018	.854	.971
VM3 (cov θ_ϵ D=2, 3=, 4=)[b]	113.04	78	.006	.841	.962
VM4 (cov θ_ϵ D=2, 3=, 4=, simplex)[c]	116.82	84	.010	.841	.964
VM5 (cov θ_ϵ D=2, 3=, 4=, simplex, β =)[d]	122.16	87	.000	.836	.962

[a] Autocorrelated residuals for Display Size 2
[b] Factor loadings for Display Sizes 3 and 4 constrained to be equal over time
[c] Simplex pattern imposed on factor structure
[d] Beta coefficients constrained equal across time
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-fit Index

as is evident in the comparison of Model VM1 with VM2 (change in $\chi^2(10, N=70) = 35.2$, $p<.01$). Model VM3 tested the hypothesis that Display Sizes 3 and 4 could be constrained equal across sessions. A comparison of Model VM3 with VM2 yielded a significant change in $\chi^2(8, N=70) = 16.07$, $p<.05$. However, this loss of fit will be accepted so that the measurement model for the VM condition will be comparable to the CM condition.

Model VM4 represents the fit of the model after imposing a first-order autoregressive process. Relative to Model VM3, there was not a significant change in $\chi^2(6, N=70) = 3.78$, $p<.80$. Thus structuring the covariances as a first-order autoregressive process does not fit the data worse than allowing them all to vary. This model is presented in Figure 19 with longitudinally standardized factor loadings. Model VM5 provides a test of the equivalence of the autoregressive paths by constraining them to be equal. Relative to Model VM4, for which the autoregressive paths are free to vary, χ^2 for Model VM5 did not increase significantly (change in $\chi^2(3, N=70) = 5.34$, $p<.20$).

Measurement Model: Variable Mapping (Session 1)

A VM performance factor was defined for each of the first and last two blocks of the first session. The results of this four-factor model are presented in Table 10 (Model VM1). Model VM1 is clearly an improvement over the null factors model. In Model VM2, the errors for Display Size 2 are correlated across sessions. Although the increase in χ^2 is not significant relative to Model VM1 (change in $\chi^2(6, N=70) = 4.89$, $p<.70$), all but one parameter estimate was significant. Consequently, the correlated errors for Display Size 2 were retained.

VM Performance - Across Practice

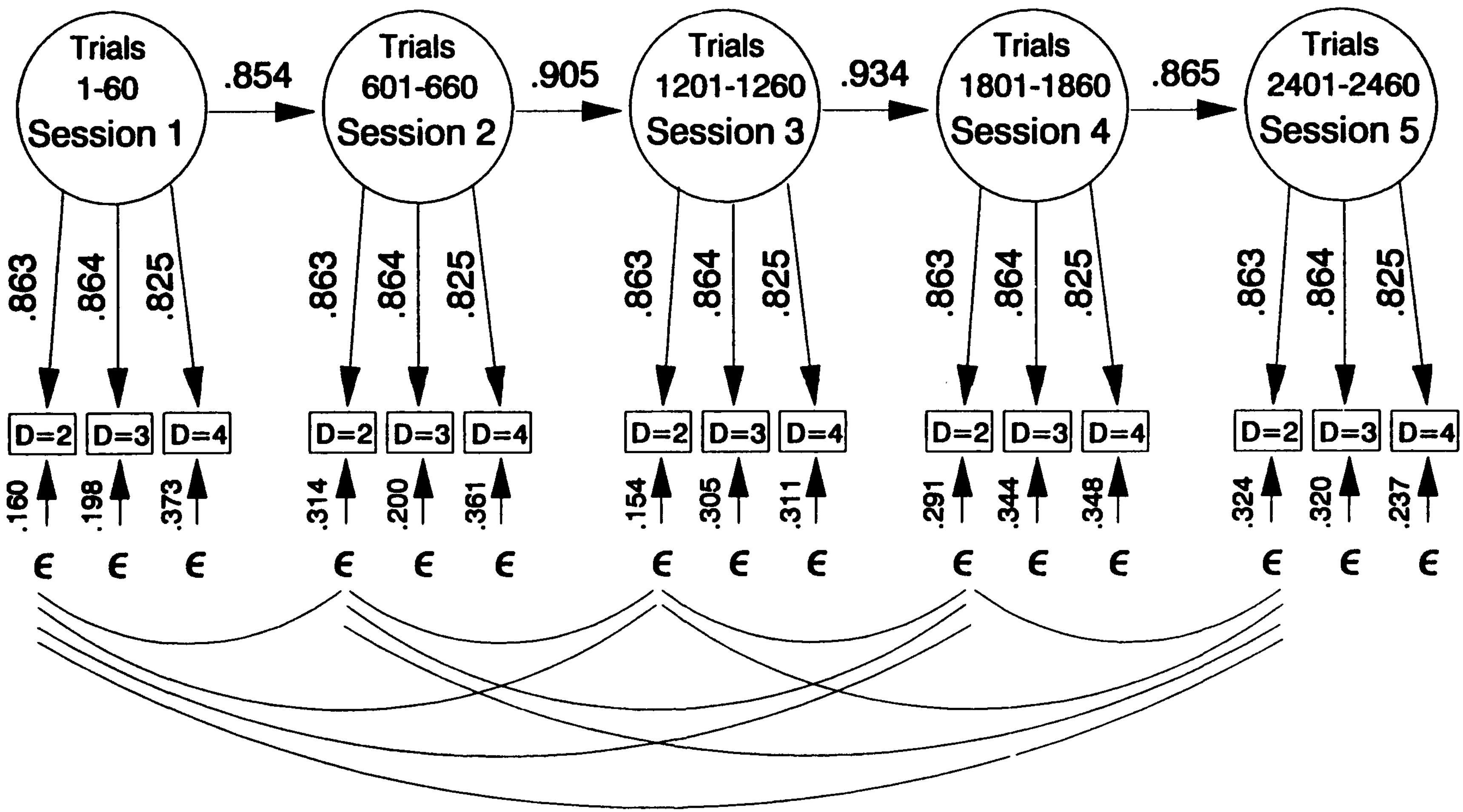


Figure 19. Measurement Model: Varied Mapping (All Sessions)

Table 10. Goodness-of-Fit Statistics: VM (Session 1)

Model	χ^2	df	p	GFI	CFI
Null Model	867.12	66	.000	.172	----
VM1	49.26	48	.423	.896	.998
VM2 (cov θ_ϵ D=2)[a]	44.37	42	.372	.909	.997
VM3 (cov θ_ϵ D=2, 3=, 4=)[b]	50.17	48	.388	.898	.997
VM4 (cov θ_ϵ D=2, 3=, 4=, simplex)[c]	53.12	51	.393	.895	.997
VM5 (cov θ_ϵ D=2, 3=, 4=, simplex, β =)[d]	53.17	52	.429	.895	.998

[a] Autocorrelated residuals for Display Size 2
[b] Factor loadings for Display Sizes 3 and 4 constrained to be equal over time
[c] Simplex pattern imposed on factor structure
[d] Beta coefficients constrained equal across time
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-fit Index

Constraining Display Sizes 3 and 4 to be equal across blocks did not worsen the fit of the model. This is evidenced by a comparison of Model VM3 with Model VM2 which yielded a nonsignificant change in $\chi^2(6, N=70) = 5.8$, $p<.50$. The autoregressive process is fit in Model VM4. Relative to Model VM3, there was not a significant change in $\chi^2(3, N=70) = 2.95$, $p<.50$. Model VM4 is represented in Figure 20 with longitudinally standardized factor loadings.

Model VM5 assessed the equality of the path coefficients for the autoregressive process. Constraining the path from Block 1 to Block 2 to be equal to the path from Block 9 to Block 10 did not yield a significant decrease in fit relative to Model VM4 (change in $\chi^2(1, N=70) = .07$, $p<.80$). Again, however, those paths will not be constrained in the structural models.

Structural Model: Consistent Mapping (All Sessions)

The best-fitting measurement model of CM search (Figure 17) was combined with the measurement model of the ability factors (Figure 16) into a structural model. The structural model assesses the relationships among the latent constructs, in this case, the ability/performance relationships.

Table 11 presents the series of models that were tested to assess the relationship between abilities and CM performance across practice sessions. The null factors model represents the model for which no factors are fit to the data. This model serves as the basis for the calculation of the CFI. Also provided for comparison is the model in which there are no paths from abilities to search but the hierarchical structure of abilities is fit along with the autoregressive process. This comparison model is referred to as the null structural model.

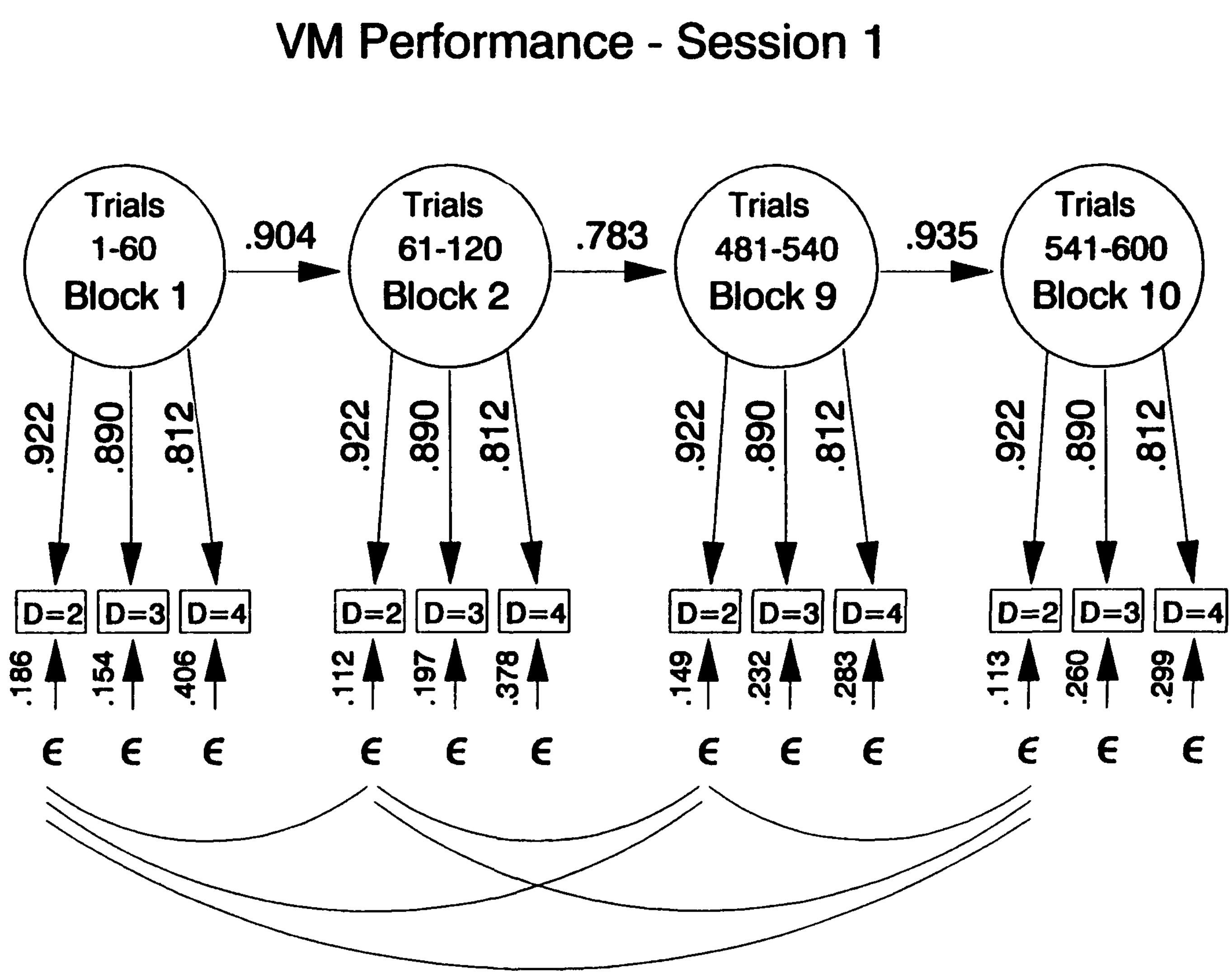


Figure 20. Measurement Model: Varied Mapping (Session 1)

Table 11. Structural Models for CM (All Sessions)

Variables in Model		χ^2	df	p	GFI	CFI
Null Factors Model		1949.29	496	.000	.190	----
Null Structural Model		582.33	445	.000	.705	.906
Influence on Session 1						
1a.	Age, Sex,	575.82	443	.000	.705	.909
1b.	Age, Sex, G	544.18	442	.001	.710	.930
1c.	Age, Sex, G, PS	540.08	441	.001	.711	.932
1d.	Age, Sex, G, PS, SMA	529.31	440	.002	.719	.938
1e.	Age, Sex, G, PS, SMA, WM	528.98	439	.002	.720	.938
1f.	Age, Sex, G, PS, SMA, WM, Gc	528.95	438	.002	.720	.937
1g.	Age, Sex, G, PS, SMA, WM, Gf	528.95	438	.002	.720	.937
1d'.	Age, Sex, G, SMA	530.40	441	.002	.720	.938
Given Age, Sex, G, SMA on Session 1, influence on Session 2:						
2a.	PS	525.49	440	.003	.720	.941
Given Age, Sex, G, SMA on Session 1, PS on Session 2, influence on Session 3:						
3a.	PS	519.05	439	.005	.725	.945

G - General Intelligence
PS - Perceptual Speed
WM - Working Memory
SMA - Semantic Memory Access
Gc - Crystallized Intelligence
Gf - Fluid Intelligence
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-Fit Index

The sequential addition of variables continued to improve the fit for models 1a through 1d. The addition of WM, Gc, and Gf did not improve the fit; therefore, direct paths from those variables were dropped from the model. Model 1d suggests that individual differences in performance on Session 1 (Trials 1-60) were influenced by individual differences in age, sex, G, PS, and SMA. However, the path coefficient from PS to initial performance was not significant and model 1d' shows that this path could be dropped from the model without loss of fit (change in $\chi^2(1, N=70) = 1.09$, $p<.30$. Thus, the critical variables for the prediction of initial-level performance were G and SMA, once age and sex differences were controlled.

Given the influence of G and SMA on initial performance, subsequent models assessed direct effects of abilities on performance in the second practice session (Trials 600-660). PS was the only variable that had an additional direct influence on Session 2 performance. Additional models were tested to assess the influence of G and SMA on Session 2 as well as on Sessions 3 through 5. None of these models yielded significant paths from G or SMA to search performance. Thus, other than the influence on Session 1, there were no other significant paths from G or SMA to search performance.

Model 3a revealed that PS also had a direct effect on Session 3 performance. Additional influences of PS on Sessions 4 and 5 were assessed in separate models but these paths were not significant. Model 3a was chosen as the final model for the following reasons. First, additional models tested whether there were other paths that were significant and/or would improve the overall fit of the model (e.g., G or SMA on Sessions 2, 3, 4, and 5; PS on Sessions 4 and 5). No significant improvements were found. Second, the modification indices in Model 3a did not reveal

any paths which, if freely estimated, might improve the fit of the model. Finally, all normalized residuals were less than 2.0 with the exception of four which were less than 3.0. This indicates that the model is reasonably well-specified (Hill, 1987).

Model 3a, presented in Figure 21, demonstrated that both G and SMA influenced initial performance, and PS had direct, additional influence on Sessions 2 and Session 3. Thus, individuals with high G and high SMA showed initially faster search; individuals with high PS ability showed greater reductions in RT with practice. Performance for the later practice sessions was well-predicted by previous performance. None of the variables had significant additional influences on these later sessions.

A model which included PS on Sessions 1, 2, and 3 was also run (χ^2 (438, N=70) = 518.21, $p<.005$, GFI=.725). The overall fit of the model was not different from Model 3a for which the path from PS to Session 1 was not included (change in $\chi^2(1, N=70) = .84$, $p<.50$). Furthermore, the path coefficients to Sessions 2 and 3 were not reduced by controlling for initial influence of PS (.216 for Session 2 and .277 for Session 3). Finally, the autoregressive coefficients remained basically unchanged from Model 3a (.677, .691, .929, .933--compare with Figure 21).

In addition to the direct effects in the final model, an inspection of the total effects (which include indirect effects) may also provide useful information (Kenny, 1979). Given that the model includes the first-order autoregressive process, the intermediate sessions of practice serve as mediators for the influences of abilities on later performance. The standardized total effects for the final model of the CM practice data are presented in Table 12. The total influence of G and SMA on CM search performance was highest for the initial trials and was reduced for later

CM Performance - Across Practice

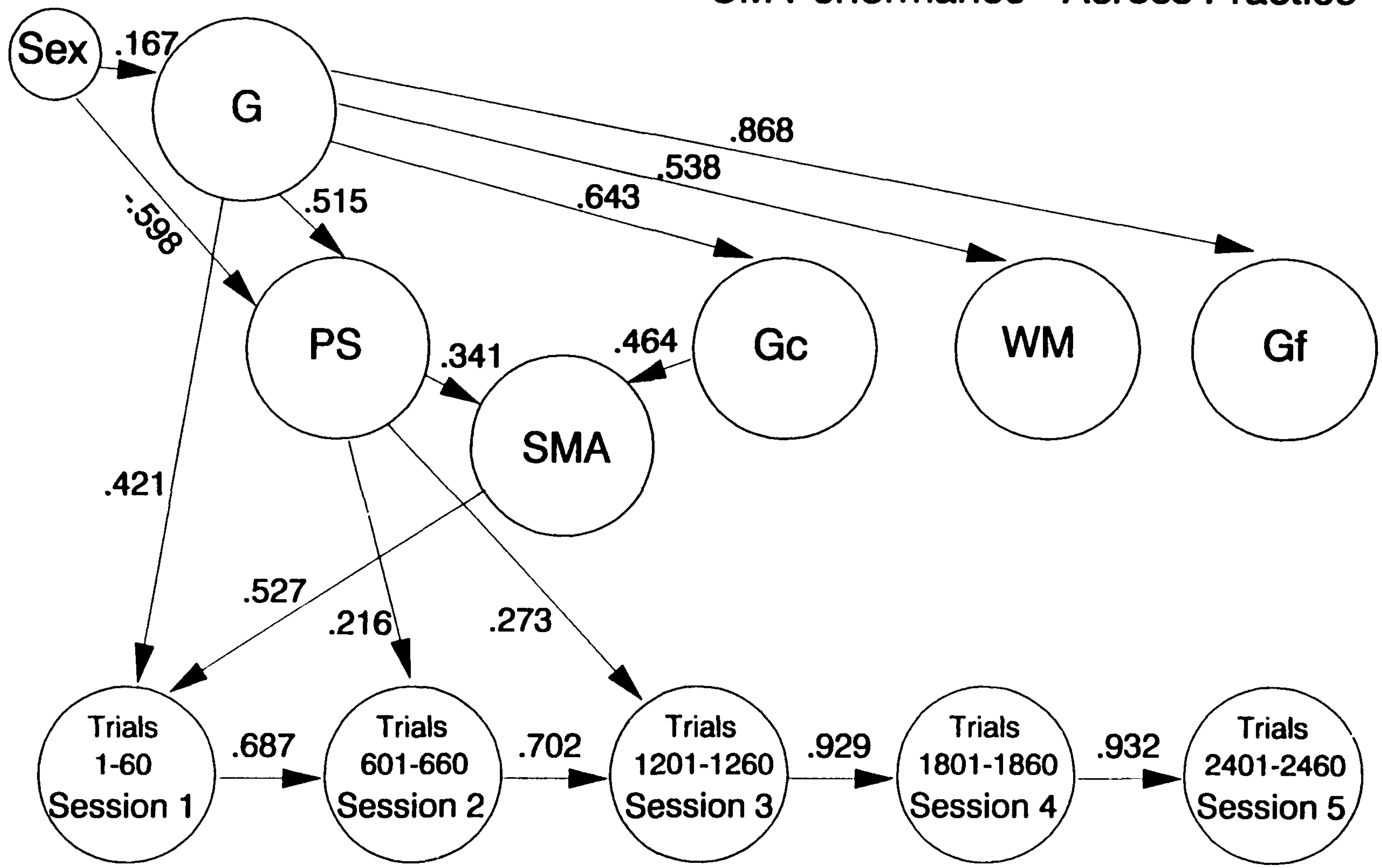


Figure 21. Structural Model: Consistent Mapping (All Sessions)

Table 12. Total Effects for CM (All Sessions)

	G	SMA	PS
Session 1			
Trials 1-60	.670	.527	.180
Session 2			
Trials 601-660	.572	.366	.340
Session 3			
Trials 1201-1260	.542	.255	.512
Session 4			
Trials 1801-1860	.503	.234	.475
Session 5			
Trials 2401-2460	.470	.217	.443

G - general intelligence
SMA - semantic memory access
PS - perceptual speed

trials. PS, on the other hand, showed an increasing, then decreasing influence on CM search performance.

Structural Model: Consistent Mapping (Session 1)

The best-fitting measurement model of CM search for Session 1 (Figure 18) was combined with the measurement model of the ability factors (Figure 16) into a structural model. χ^2 statistics for the ability/performance models for the first CM practice session are presented in Table 13 along with the null factors and null structural models. The pattern of influence for the first block of trials is necessarily the same as that reported above because it is the same data for the same subjects. Thus, age, sex, G, and SMA all influence the first block of CM search. The value of χ^2 and the degrees of freedom differ because Blocks 1, 2, 9, and 10 are included from Session 1 (whereas in the previous models the first block of each session was included). The major focus of interest for this series of models was whether any of the ability factors directly influenced these earlier blocks.

The influence of SMA on each block was assessed; none of the paths were significant other than the influence on Block 1. Similarly, the influence of PS on each block was assessed; only the influence on Block 1 was significant. The influence of G on each block was also assessed. Again, the only significant path was from G to Block 1. The lack of additional influences led to the acceptance of 1d' as the best-fitting model for this data. This model is presented in Figure 22.

As is evident in Figure 22, the coefficient for the path from Block 2 to Block 9 is fairly low (.778). This might indicate the need for an additional influence of one of the abilities on Block 9. However, the seven-block gap between Blocks 2 and 9 might account for the low

Table 13. Structural Models for CM (Session 1)

Variables in Model		χ^2	df	p	GFI	CFI
Null Factors Model		1626.56	406	.000	.222	----
Null Structural Model		482.87	362	.000	.720	.901
Influence on Block 1						
1a.	Age, Sex,	475.99	360	.000	.719	.905
1b.	Age, Sex, G	444.95	359	.001	.727	.930
1c.	Age, Sex, G, PS	441.42	358	.002	.729	.932
1d.	Age, Sex, G, PS, SMA	431.25	357	.004	.736	.939
1e.	Age, Sex, G, PS, SMA, WM	430.27	356	.004	.736	.939
1f.	Age, Sex, G, PS, SMA, WM, Gc	430.26	355	.004	.736	.938
1g.	Age, Sex, G, PS, SMA, WM, Gf	431.96	355	.004	.736	.937
1d'.	Age, Sex, G, SMA	431.96	358	.004	.736	.939

G - General Intelligence
PS - Perceptual Speed
WM - Working Memory
SMA - Semantic Memory Access
Gc - Crystallized Intelligence
Gf - Fluid Intelligence
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-Fit Index

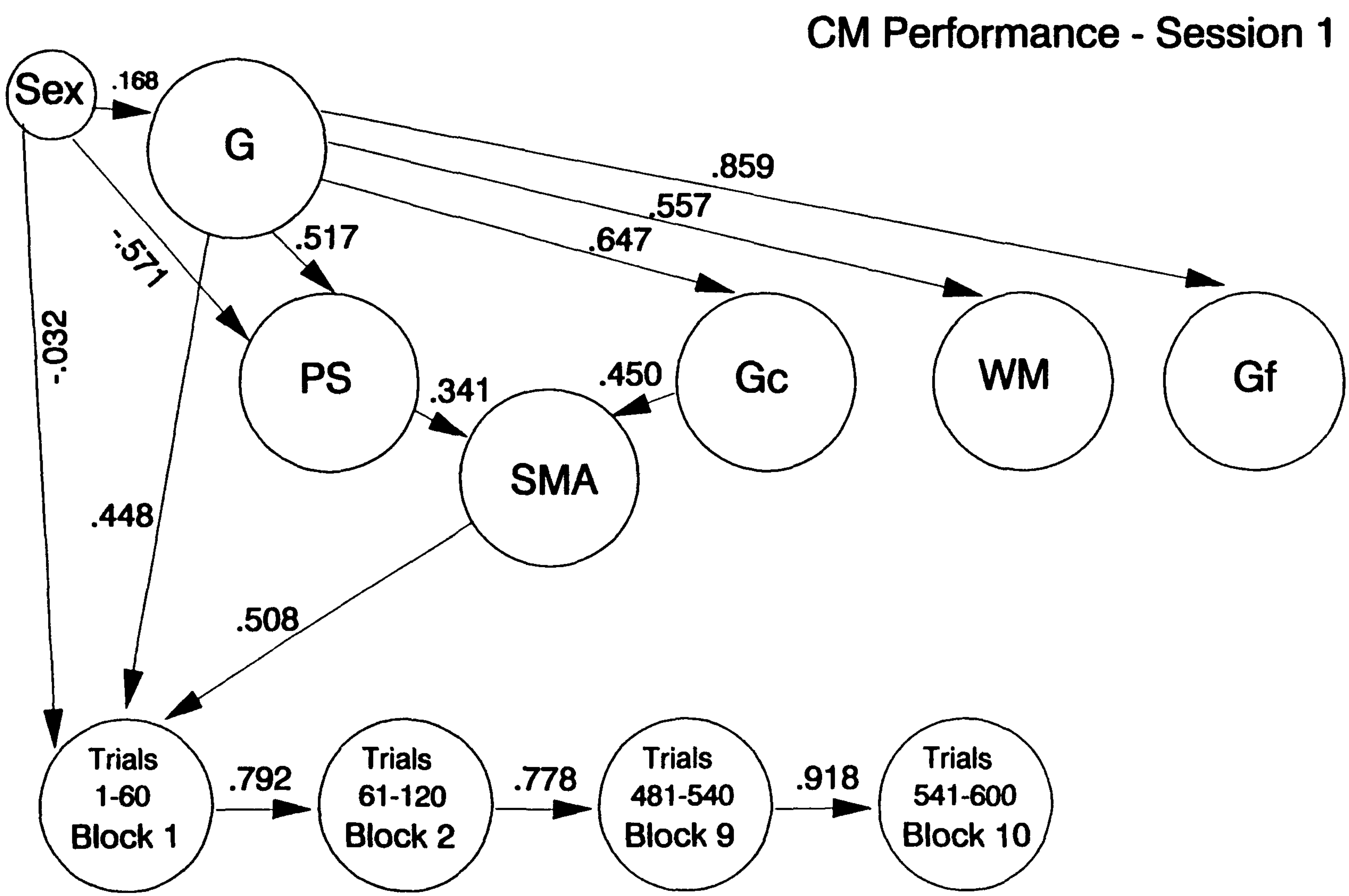


Figure 22. Structural Model: Consistent Mapping (Session 1)

coefficient. Suppose, for example that there were a path coefficient between each block of .965. When taken to the seventh power, the value would be .779. Thus it is important to note that a high relationship between blocks in this first session may be masked because the intervening blocks are not included in the model. There was a marginally significant path coefficient from PS to Block 2 of .207. However, including this path did not significantly improve the overall fit of the model nor did it result in an increase in the stability coefficients between blocks. The autoregressive coefficients were, in order, .709, .780, and .918. Furthermore, allowing additional paths from SMA, PS, or G to Block 9 did not increase the coefficient from Block 2 to Block 9. For example, a model with an estimated path from PS to Block 9 yielded a nonsignificant path coefficient of .160, and the autoregressive coefficients were, in order, .791, .717, and .916.

The total effects for model 1d' of the first CM practice session are presented in Table 14. The total influence of G on CM search performance was highest for the initial trials and somewhat reduced for later trials. SMA also had a significant influence on initial performance but a greatly reduced influence on later sessions. PS did not heavily influence CM search performance.

Discussion: Ability/Performance Relationships for Consistent Mapping

As was predicted based on previous investigations, G had a significant influence on initial-level performance. This influence was reduced as a function of practice (e.g., Ackerman, 1988; Fleishman, 1972). Although there was no additional influence of G on the later practice sessions, the total effect of G on performance remained fairly high (Table 12).

Table 14. Total Effects for CM (Session 1)

	G	SMA	PS
Block 1 Trials 1-60	.686	.506	.173
Block 2 Trials 61-120	.543	.404	.137
Block 9 Trials 481-540	.423	.313	.107
Block 10 Trials 541-600	.388	.289	.098

G - general intelligence
SMA - semantic memory access
PS - perceptual speed

In addition to G, SMA also had a significant influence on initial performance. This ability/performance relationship has not previously been reported for search tasks because SMA was not separately assessed (cf. Ackerman, 1986, 1988). However, the influence of SMA on category search performance is not surprising. Recall that an important component of category search involves the strength of the association between exemplars and their higher-order categories. Thus, individuals with faster SMA will perform better on the search task. This influence of SMA on search performance drops out within the first session (Table 12). This most likely occurs because the categories used in the experiment are well-known and any differential advantage for subjects with faster SMA is eliminated as the exemplar-to-category link is strengthened for all subjects. This strengthening can take place quickly in the CM condition because there is only one CM target category for each subject. Note that the finding of an influence of SMA on early performance differs from Kyllonen and Woltz's (1989) contention that the speed with which items can be retrieved from long-term memory (LTM) will be most important for differentiating between individuals late in practice.

A simple first-order autoregressive process did not provide the best fit to the data. That is, it was not sufficient to predict initial-level performance based on background abilities, then predict later performance solely from previous performance via autoregression. As is evident in Figure 21, PS had additional influences on Sessions 2 and 3. These additional paths suggest that there are individual differences in the magnitude of learning (i.e., individuals are changing their rank orderings across sessions) and these individual differences can be predicted by individual differences in PS. Individuals with greater PS ability improve faster. Such individual differences in learning rate may account for the additional influence of PS

on Session 3. Even after 1200 CM practice trials, individual differences in PS have a direct effect on individual differences in search performance after controlling for prior levels of search performance. The PS/performance relationship does not follow Ackerman's prediction that the influence of PS should be greatly reduced with practice. It is of course possible that 1200 practice trials were not sufficient to eliminate the PS influence. However, Table 12 reveals that the total effect of PS (mediated through Session 4) remained high even after 2400 trials (.443), although it is slightly reduced relative to performance at 1200 trials (.512). It seems likely, then, that even for well-practiced individuals, PS ability is a differentiating factor in determining successful CM performance for the present class of tasks.

By Session 4 (1800 practice trials), CM performance was well-predicted by previous-level performance and there were no additional ability/performance paths. Thus, although mean RTs continued to decrease beyond this point (Figure 4), the relative positions of individuals remained stable.

Prediction of initial performance with G and SMA along with a first-order autoregressive process provided the best fit to the data for performance in Session 1. Thus, early performance is primarily a function of G and SMA. There were no direct paths from abilities to Block 9 performance. This is surprising in light of the relatively low stability coefficients between blocks. However, it was suggested that the low stability coefficients may be artifactual due to the gap between Blocks 2 and 9. However, by Session 2, the influence of G and SMA was reduced whereas the influence of PS was increased. The lack of additional ability/performance relationships for the later blocks of Session 1 along with the presence of a PS path for Session 2 is intriguing. One possible explanation for this pattern is

that there was some "consolidation" across sessions that covaries with individual differences in PS. Another possible explanation might be that fatigue at the end of Session 1 masked the emerging relationship to PS. Additionally, the path from PS to Block 3 may have been only marginally significant due to low power; if so, this path might be important but its significance could not be detected in the present sample.

While this pattern of results is comparable to the general pattern predicted by Ackerman, his results revealed a relationship between PS and hybrid memory/visual search performance in less than 200 trials. Perhaps the more complex visual environment (i.e., a larger display) of the current, "pure" visual-search task had an effect on the rate at which the influence of G drops out and PS becomes more important for differentiating individuals in terms of their search performance.

To summarize, the general pattern of the ability/performance relationship is consistent with the general predictions: G and SMA are influential in determining initial-level search performance and PS is more influential for later performance. A contrary finding was that SRT (i.e., PM) was not related to performance even after 3000 practice trials. The lack of a relationship between PM and search was not specific to SRT because the zero-order correlations were low for the paper and pencil measures of PM as well (i.e., Making Xs and Crossing Lines).

Structural Model: Variable Mapping (All Sessions)

The best-fitting measurement model of VM search (Figure 19) was combined with the measurement model of the ability factors (Figure 16) into a structural model. χ^2 statistics for the ability/performance models for the VM search condition across sessions are presented in Table 15.

Table 15. Structural Models for VM (All Sessions)

Variables in Model		χ^2	df	p	GFI	CFI
Null Factors Model		1961.42	496	.000	.182	----
Null Structural Model		659.63	445	.000	.699	.854
Influence on Session 1						
1a.	Age, Sex,	657.76	443	.000	.698	.853
1b.	Age, Sex, G	632.09	442	.000	.701	.870
1c.	Age, Sex, G, PS	626.86	441	.000	.702	.873
1d.	Age, Sex, G, PS, SMA	600.61	440	.000	.715	.890
1e.	Age, Sex, G, PS, SMA, WM	600.42	439	.000	.716	.890
1f.	Age, Sex, G, PS, SMA, WM, Gc	597.57	438	.000	.716	.891
1d'.	Age, Sex, G, SMA	600.62	441	.000	.715	.891
Given Age, Sex, G, SMA on Session 1, influence on Session 2:						
2a.	PS	595.77	440	.000	.717	.891

G - General Intelligence
WM - Working Memory
Gc - Crystallized Intelligence
PS - Perceptual Speed
SMA - Semantic Memory Access
Gf - Fluid Intelligence
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-Fit Index

The pattern of influence for the first block of trials is similar to that observed for CM; that is, G and SMA have significant influence on early performance, given also control for age and sex differences. An investigation of the influences of G, PS, and SMA on Session 2 revealed that only PS had a significant additional influence on Session 2 (Model 2a). The influences of G, PS, and SMA were also assessed for later sessions of performance, but none of these paths were significant nor were there significant improvements in the fit of these models relative to Model 2a. Performance for Sessions 3, 4, and 5 was well-predicted by previous-level performance. Thus, the best fitting structural model was 2a; this model is presented in Figure 23.

Total effects for VM practice data are presented in Table 16. The total influence of G and SMA on VM search performance was highest for the initial trials and reduced for later trials. PS, on the other hand, showed an increasing, then decreasing influence on VM search performance.

Structural Model: Variable Mapping (Session 1)

The best-fitting measurement model of VM search for Session 1 (Figure 20) was combined with the measurement model of the ability factors (Figure 16) into a structural model to assess ability/performance relationships for the first session of VM performance. The results of these models are presented in Table 17. The pattern of influence on Block 1 was necessarily the same as reported above for Session 1 because it is the same data. Thus, age, sex, G, and SMA influenced Block 1 performance. Again, the focus of this analysis was to determine if there were additional influences of abilities on performance in the first session of performance. Additional effects of G, SMA, and PS on Blocks 2, 9, and 10 were assessed in separate models. There

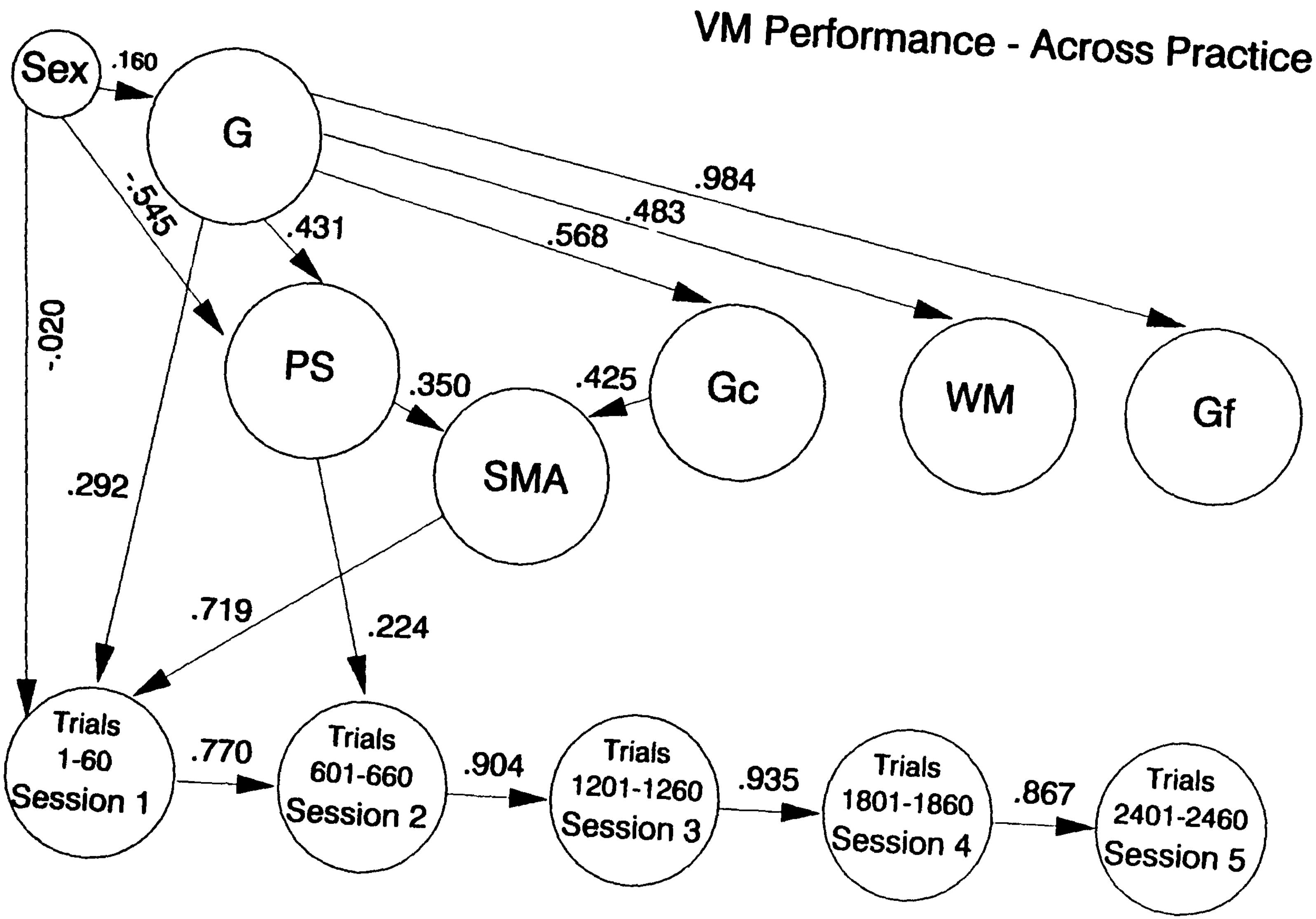


Figure 23. Structural Model: Varied Mapping (All Sessions)

Table 16. Total Effects for VM (All Sessions)

	G	SMA	PS
Session 1 Trials 1-60	.549	.725	.252
Session 2 Trials 601-660	.519	.551	.417
Session 3 Trials 1201-1260	.469	.501	.377
Session 4 Trials 1801-1860	.439	.470	.353
Session 5 Trials 2401-2460	.380	.405	.306

G - general intelligence
SMA - semantic memory access
PS - perceptual speed

Table 17. Structural Models for VM (Session 1)

Variables in Model		χ^2	df	p	GFI	CFI
Null Factors Model		1722.26	406	.000	.202	----
Null Structural Model		519.22	361	.000	.717	.880
Influence on Block 1						
1a.	Age, Sex,	516.92	359	.000	.716	.880
1b.	Age, Sex, G	489.93	358	.000	.715	.900
1c.	Age, Sex, G, PS	483.77	357	.000	.719	.904
1d.	Age, Sex, G, PS, SMA	460.75	356	.000	.734	.920
1e.	Age, Sex, G, PS, SMA, WM	460.65	355	.000	.734	.920
1f.	Age, Sex, G, PS, SMA, WM, Gc	458.32	354	.000	.736	.921
1d'.	Age, Sex, G, SMA	460.99	357	.000	.734	.921

G - General Intelligence
PS - Perceptual Speed
WM - Working Memory
SMA - Semantic Memory Access
Gc - Crystallized Intelligence
Gf - Fluid Intelligence
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-Fit Index

were no significant paths from any of those abilities to the remaining blocks of Session 1 (i.e., beyond Block 1). Thus, the best-fitting model for Session 1 of VM practice is 1d' (Figure 24).

The total effects for this model of the first VM practice session are presented in Table 18. The total influence of G on VM search performance was highest for the initial trials and somewhat reduced for later trials. The influence of SMA was also largest for the first session and somewhat reduced for later sessions. PS, on the other hand, showed a small, nonsignificant influence on VM search performance.

Discussion: Ability/Performance Relationships for Variable Mapping

G and SMA both influenced initial-level performance. Table 16 reveals that the total effects of both were reduced with practice but remained fairly high. As with the CM data, it was not sufficient to simply predict initial performance, then utilize a first-order autoregressive process to predict later performance. PS had an additional influence on VM performance in Session 2. Subsequent sessions (3 through 5) were well-predicted by prior performance. Thus, by Session 3, performance on the VM task had stabilized both in terms of mean RT (Figure 4) and covariance structure. As with the CM condition, analyzing performance within the first session of practice did not reveal any additional ability/performance relationships. The implications of the ability/performance relationships are discussed in the following section in contrast to the ability/performance relations for the CM condition.

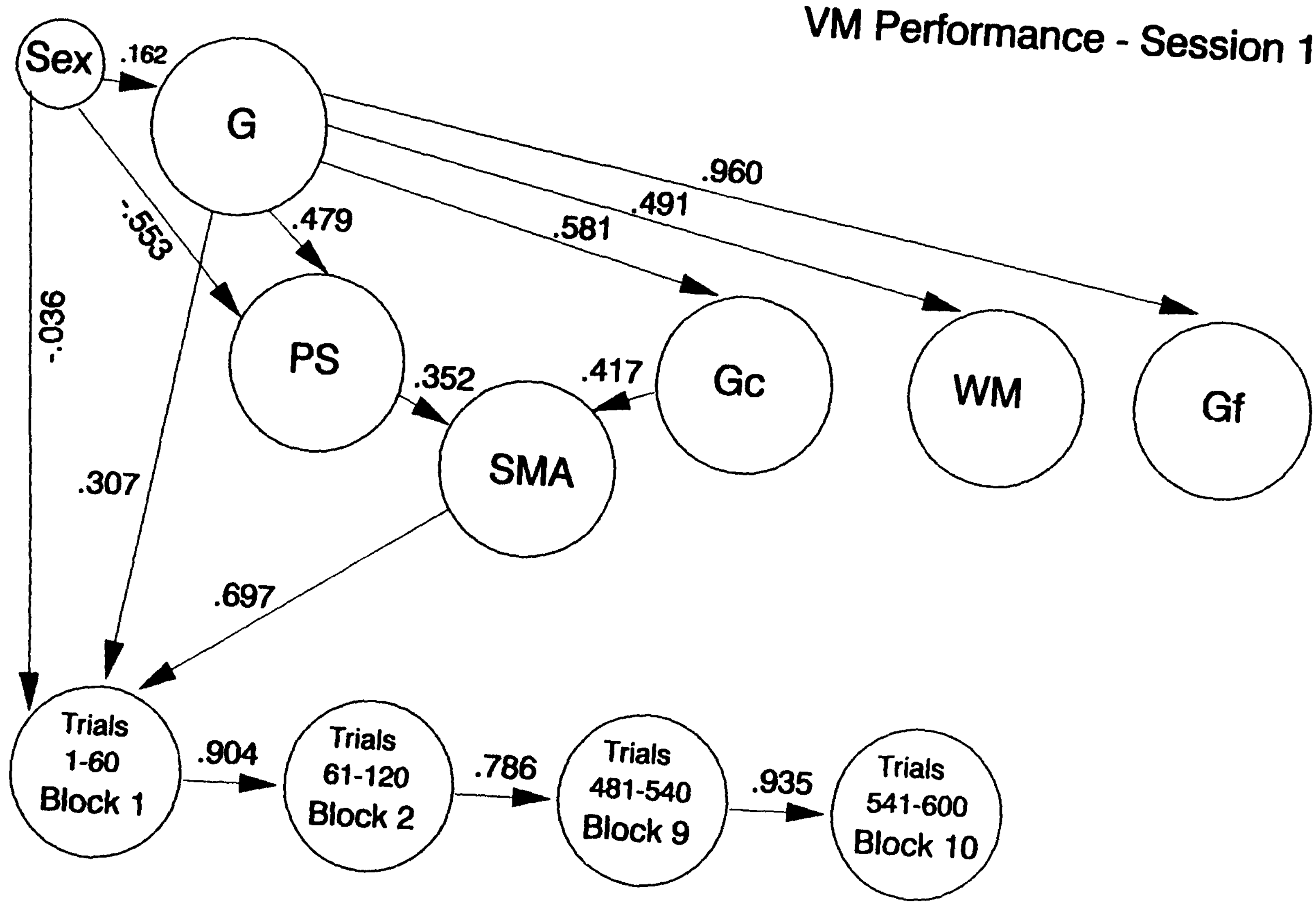


Figure 24. Structural Model: Varied Mapping (Session 1)

Table 18. Total Effects for VM (Session 1)

	G	SMA	PS
Block 1 Trials 1-60	.591	.698	.245
Block 2 Trials 61-120	.534	.632	.222
Block 9 Trials 481-540	.420	.494	.174
Block 10 Trials 541-600	.393	.462	.163

G - general intelligence
SMA - semantic memory access
PS - perceptual speed

Consistent/Variable Mapping Comparisons

Somewhat surprisingly, the ability/performance relationships are very similar for CM and VM (compare Figures 21 and 23). Both practice conditions were primarily influenced by G and SMA in the first session, and by PS in the second session. However, CM performance had an additional direct influence of PS on Session 3 whereas VM performance did not. Thus, in the CM condition, the relative position of individuals is a stronger function of PS ability across more trials of practice. In conjunction with the normative data, this finding suggests that performance continues to improve with CM practice and the improvement is a function of the ability to make rapid comparisons and responses. PS may be an important factor in the development of optimal search strategies as well as an automatic response. Also, the additional influence of PS on Session 3 might be an indication that automatic response development does not occur at a constant rate for all individuals. In the CM condition, performance improvements are still occurring in Session 3, but have stabilized by then in the VM condition. This is consistent with the mean RT results.

A comparison of Tables 12 and 16 reveals that while the patterns for G and SMA are similar, the magnitude of the SMA influence differs across CM and VM. The LISREL estimates (i.e., the unstandardized coefficients) and standard errors for the final ability/performance models for the CM and VM conditions are presented in Table 19. The influence of G on performance was higher for initial CM performance than initial VM performance, but this difference is not significant ($z=.596$, $p<.28$).[11] However, the trend may be a

[11]This is a conservative test of the difference between the parameters because it does not account for the covariance between the standard errors due to the within-subject design.

Table 19. LISREL Estimates for CM and VM

	LISREL Estimates (standard errors)		
CM - All Sessions	**G**	**SMA**	**PS**
Age, Sex, G, SMA on Session 1	1.069 (0.461)	0.103 (0.027)	-------------
PS on Session 2	-------------	-------------	0.346 (0.185)
PS on Session 3	-------------	-------------	0.386 (0.158)
VM - All Sessions	**G**	**SMA**	**PS**
Age, Sex, G, SMA on Session 1,	0.712 (0.383)	0.160 (0.029)	-------------
PS on Session 2	-------------	-------------	0.511 (0.235)
CM - Session 1	**G**	**SMA**	**PS**
Age, Sex, G, SMA on Block 1	1.086 (0.441)	0.093 (0.025)	-------------
VM - Session 1	**G**	**SMA**	**PS**
Age, Sex, G, SMA on Block 1	0.850 (0.415)	0.158 (0.030)	-------------

function of the fact that the influence of G has larger initial impact on very early performance, and the first block of the CM condition was prior to the first block of the VM condition. This is consistent with Ackerman's (1988) data in which the G/performance correlation was reduced within the first 60 practice trials.

A comparison of the path coefficient for SMA to initial performance reveals the opposite trend: there is a higher relationship between SMA and VM performance than between SMA and CM performance. The difference is not significant with an alpha set at .05, $z=1.439$, $p<.075$; however, given the conservative test it may in fact be a meaningful difference. There are two potential explanations for the CM/VM difference in SMA influence. First, it could be that the higher influence of G on early CM performance overshadows the influence of SMA. That is, the ability to learn how to perform the task in a more general sense is most predictive of initial performance. Because VM is the second block of practice, the influence of G is reduced and SMA has a higher relationship to performance. Although this possibility cannot be ruled out, it seems unlikely. A more compelling explanation is related to the fact that there are five potential target categories in the VM condition but only one target category in the CM condition. Hence, SMA is a more important component of the VM task. Although the actual values are not comparable in Tables 12 and 16, the trends may be compared. In the CM condition, the influence of SMA is greatly reduced by Session 2 and is consistently less than the influence of G. In the VM condition, the influence of SMA is also reduced by Session 2 but it is consistently higher than the influence of G.

Measurement Model: Transfer Session

In the transfer session, subjects received two new conditions: 1) CM Reversal - a condition in which the roles

of the previous CM targets and distractors were reversed, and 2) New CM - a condition in which two of the previous VM categories were paired to form a CM condition. The correlations between performance in the first block of each transfer condition, Early CM, Late CM, and Late VM are provided in Table 20. The correlations with Late VM are provided for comparison purposes, but only the four CM conditions were included in the measurement model.

A four-factor model was fit to the data. Early CM, Late CM, CM Reversal, and New CM were each defined using the respective mean RTs for Display Sizes 2, 3, and 4. The metric of the factor was defined by fixing to 1.0 the loading of Display Size 2. The fit for this four-factor model, denoted Trans1, is presented in Table 21 along with the null factors model. Model Trans2 tested the hypothesis that the factor loadings for Display Sizes 3 and 4 could be constrained equal across conditions without a significant loss of fit. A comparison of Model Trans2 with Trans1 supports this hypothesis with a nonsignificant change in $\chi^2(3, N=70) = 4.69$, $\underline{p}<.20$.

Two variants of a first-order autoregressive process were compared. In Model Trans3, early CM predicts Late CM, CM Reversal, and New CM. This model did not fit the data very well and was significantly worse than Model Trans2 (change in $\chi^2(3, N=70) = 17.46$, $\underline{p}<.01$). In Model Trans4, Early CM predicted Late CM, but Late CM served as the predictor for CM Reversal and New CM. However, this model did not fit the data very well either; it was also significantly worse than Model Trans2 (change in $\chi^2(3, N=70) = 17.39$, $\underline{p}<.01$).

Model Trans5 retained the autoregressive process from Model Trans4 but also allowed the covariance between the CM Reversal and New CM conditions to be freely estimated. This represents the within-session covariance that cannot be

Table 20. Training and Transfer Correlation Matrix[a]

	1	2	3	4	5
1. Early CM	1.0				
2. Late CM	.515	1.0			
3. Late VM	.447	.507	1.0		
4. CM Reversal	.445	.441	.565	1.0	
5. New CM	.373	.435	.446	.571	1.0
Mean	871.8	557.9	738.7	780.1	713.3
SD	119.9	66.9	77.7	95.3	86.6

[a] Correlations > .23 are significant at .05

Table 21. Goodness-of-Fit Statistics: Transfer Session

Model	χ^2	df	p	GFI	CFI
Null Model	633.69	66	.000	.283	----
TRANS1	55.35	48	.217	.886	.987
TRANS2 (3=, 4=)[a]	60.04	54	.266	.876	.989
TRANS3 (3=, 4=, Early CM)[b]	77.50	57	.037	.837	.964
TRANS4 (3=, 4=, Late CM)[c]	77.43	57	.037	.847	.964
TRANS5 (3=, 4=, Late CM, REV-NCM)[d]	64.96	56	.193	.871	.984

[a] Factor loadings for Display Sizes 3 and 4 constrained to be equal over time.
[b] Early CM predicts Late CM, CM Reversal, and New CM.
[c] Early CM predicts Late CM, Late CM predicts CM Reversal and New CM.
[d] Early CM predicts Late CM, Late CM predicts CM Reversal and New CM, covariance between CM Reversal and New CM freely estimated.
GFI - LISREL Goodness-of-Fit Index
CFI - Bentler (1990) Comparative-Fit Index

accounted for by an autoregressive process. This significantly improved the fit of the model (change in $\chi^2(1, N=70) = 12.46$, $p<.01$). The fit of a second-order autoregressive process (i.e., an additional path from Early CM) was assessed both for CM Reversal and New CM but neither of these paths were significant nor was there a significant improvement in the overall fit of the model.

Relative to Model Trans2, which freely estimated the covariances among the conditions, Model Trans5 did not provide a worse fit (change in $\chi^2(2, N=70) = 4.92$, $p<.10$). Thus, the first-order autoregressive process, along with a freely estimated covariance between the two transfer conditions, provided as good a fit as the model with all free covariances. Model Trans5, presented in Figure 25, was thus retained as the best-fitting model.

Structural Model: Transfer Session

Based on the accepted structural model of the CM performance data (Figure 21), an initial model was assessed in which Early CM performance was predicted by age, sex, G, and SMA; and Late CM performance had an additional influence of PS. (Note that the path from PS to Late CM was added because the intermediate sessions were not included in this model.) The χ^2 statistic for this model is presented in Table 22 and is denoted 1a. In addition, the paths in Figure 25 were retained. Early CM predicted Late CM, Late CM predicted both New CM and CM Reversal and the covariance between the two transfer conditions was freely estimated. This model provided only a reasonable fit to the data. Recall that the mean RT data demonstrated that performance was disrupted in the two transfer conditions. Additional models were run to assess whether the abilities that were important for initial-level performance might also be important predictors for the New CM and CM Reversal conditions. First, a model was tested which included a path

Transfer Data

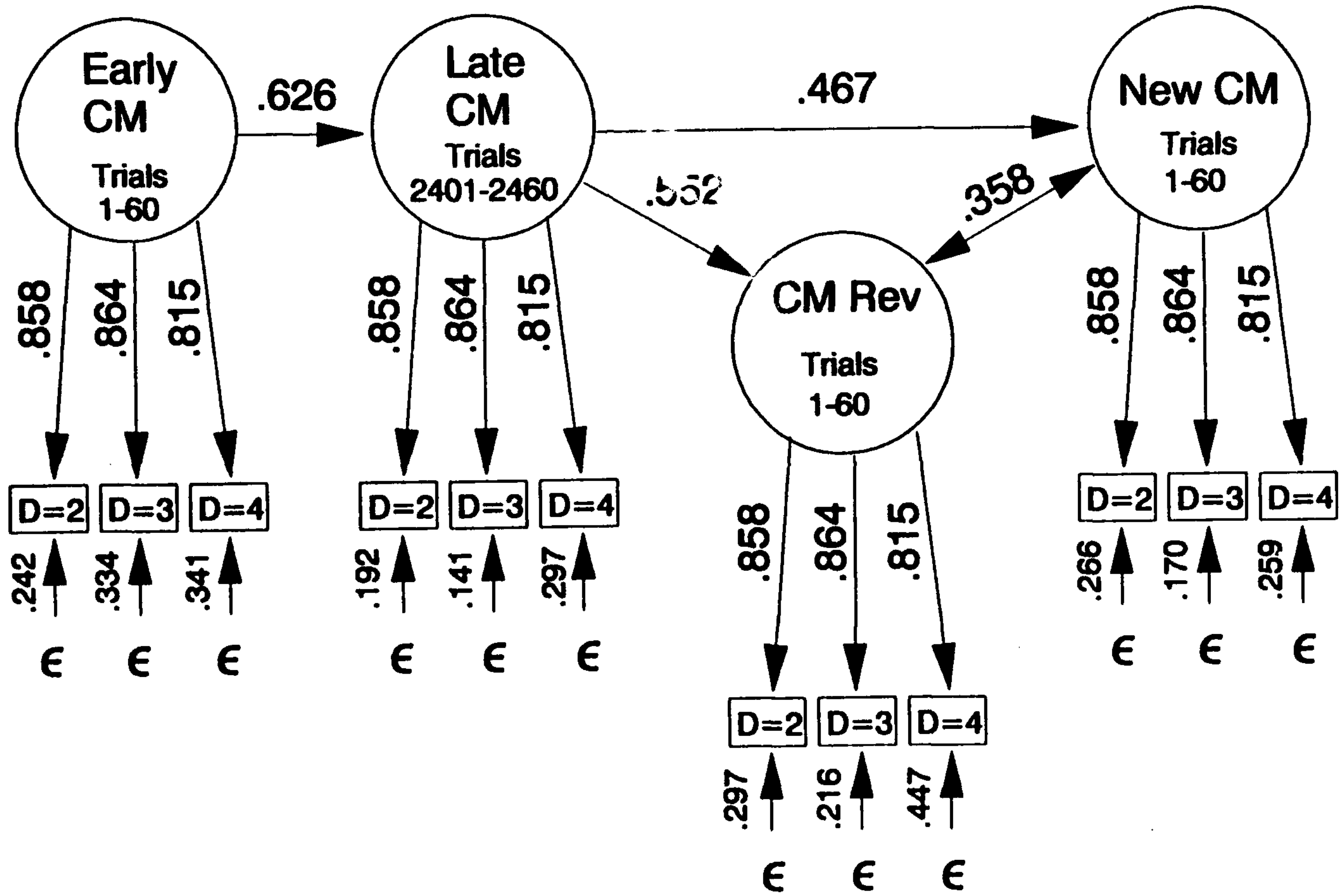


Figure 25. Measurement Model: Transfer Session

Table 22. Structural Models for Transfer Session

Variables in Model	χ^2	df	p	GFI	CFI
Null Factors Model	1482.60	406	.000	.263	----
Null Structural Model	524.10	366	.000	.708	.853
1a. Age, Sex, G, SMA --> Early CM					
PS --> Late CM					
Early CM --> Late CM					
Late CM --> CM Rev					
Late CM --> New CM					
CM Rev <--> New CM	461.80	361	.000	.729	.906
1b. Age, Sex, G, SMA --> Early CM					
PS --> Late CM					
Early CM --> Late CM					
Late CM --> CM Rev					
Late CM --> New CM					
CM Rev <--> New CM					
PS, SMA --> CM Rev					
PS, SMA --> New CM	435.32	357	.003	.746	.927
1c. Age, Sex, G, SMA --> Early CM					
PS --> Late CM					
Early CM --> Late CM					
Late CM --> CM Rev					
Late CM --> New CM					
PS, SMA --> CM Rev					
PS, SMA --> New CM	436.83	358	.003	.744	.927

G - General Intelligence
PS - Perceptual Speed
WM - Working Memory
SMA - Semantic Memory Access
Gc - Crystallized Intelligence
Gf - Fluid Intelligence
SRT - Simple Reaction Time
GFI - LISREL Goodness-of-Fit-Index
CFI - Bentler (1990) Comparative-Fit Index

from G to CM Reversal and New CM. Neither of these path coefficients were significant. Additional models tested whether PS and/or SMA had significant influence on transfer performance. In fact, the path coefficients for both PS and SMA to CM Reversal and New CM were significant. An inspection of the modification indices did not reveal any other potential ability/performance relationships. In Model 1b, PS and SMA influenced New CM and CM Reversal. This model is a significant improvement over Model 1a which did not estimate any paths between abilities and the transfer conditions (change in $\chi^2(4, N=70) = 26.48$, $p<.01$). Model 1c tested the possibility that the covariance between CM Reversal and New CM could be fixed to zero without a loss of fit. This covariance was estimated in the measurement model to improve the fit of the model, but the inclusion of the PS and SMA paths to the transfer conditions may have eliminated its utility. A comparison of Model 1c with Model 1b supports this proposal with a nonsignificant change in $\chi^2(1, N=70) = 1.51$, $p<.70$. Model 1c was chosen as the best-fitting model and is presented in Figure 26. The total effects of abilities on performance in this model are presented in Table 23. Note that the relative influence of SMA and PS on performance is similar for the CM Reversal and the New CM conditions.

Discussion: Transfer Data

The ability/performance relationships for the transfer session demonstrate that performance is not completely predicted either by final-level CM performance or by initial CM performance. Instead, additional influences of PS and SMA are necessary to improve the fit of the model to the data. Thus, successful performance under transfer conditions requires PS and SMA to a greater degree than for the "automated" search in the CM condition. Interestingly, PS has a strong influence on transfer performance which was

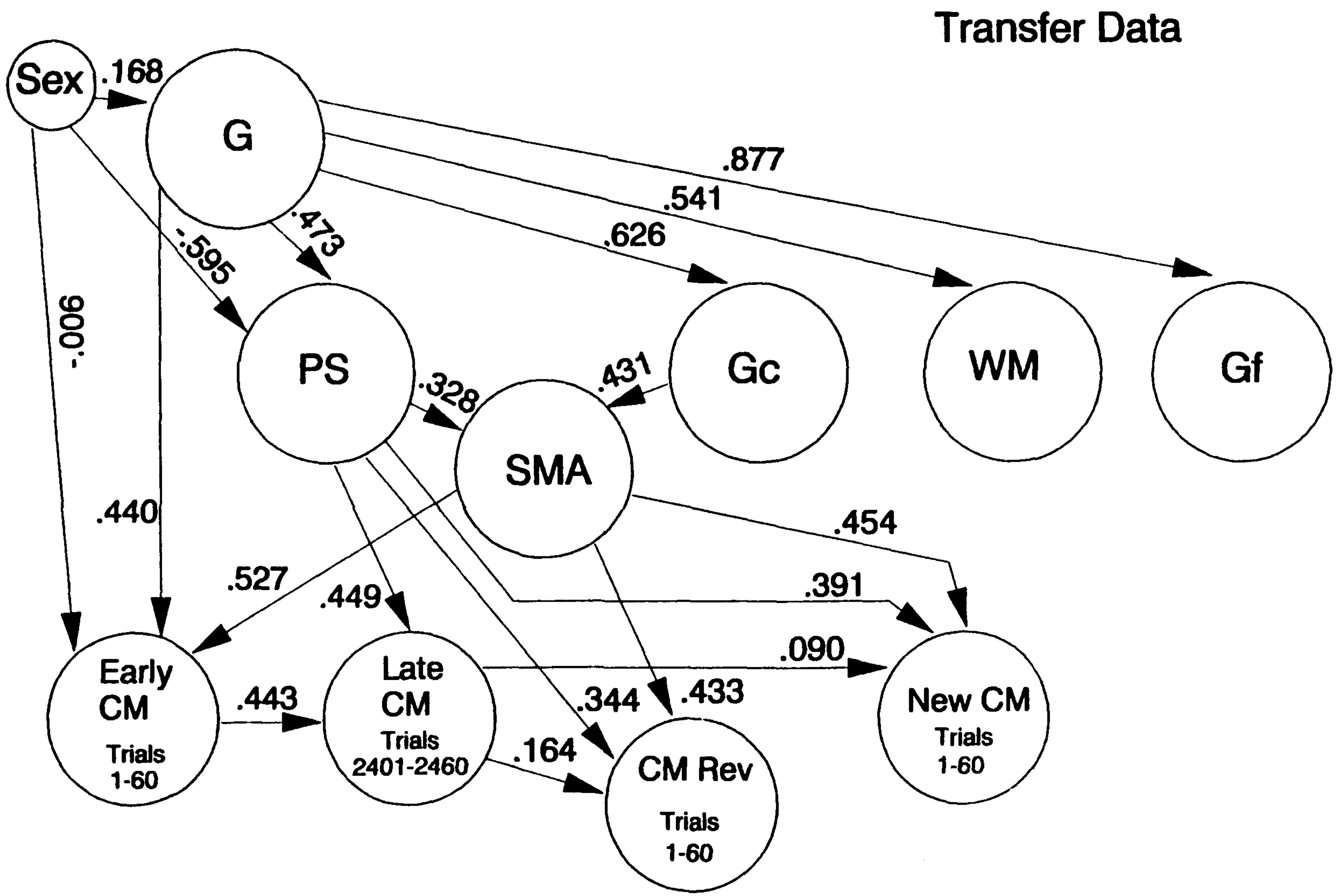


Figure 26. Structural Model: Transfer Session

Table 23. Total Effects: Transfer Session

	G	SMA	PS
Early CM	.665	.525	.173
Late CM	.500	.240	.486
CM Reversal	.430	.470	.566
New CM	.393	.476	.518

G - general intelligence
SMA - semantic memory access
PS - perceptual speed

not the case for initial CM performance. Subjects may have benefitted from general search practice; thus, G was less influential on transfer performance. If final-level CM performance were dominated by an automatic response (which is necessarily stimulus-specific), subjects would have a disadvantage in the two transfer conditions. As a result, performance is well-predicted by PS and SMA abilities rather than previous CM performance alone.

Note that the ability/performance functions do not differ for the two transfer conditions. Thus, changing the category pairings either by combining two VM categories or reversing the roles of previous CM targets and distractors disrupts performance such that PS and SMA abilities are related to individual differences in performance. The structural models only compare shifts in the covariance structures, but the mean RT data demonstrated a differential amount of disruption for the New CM and CM Reversal conditions. The lack of difference in the ability/performance relationships, along with the large difference in the mean RT disruption for the two transfer conditions, is important. The normative and individual differences analyses would result in different conclusions if either were viewed in isolation. However, taken together, the results provide important information about mean differences in transfer as well as the processes underlying successful performance.

LIMITATIONS/QUALIFICATIONS/CAVEATS

Potential Problems with Sample Size. The present study included 70 subjects--more than enough for the normative analyses conducted on the mean-level data. However, in terms of the modeling analysis, 70 subjects per group is considered a small-sample analysis. Bollen (1989) reviewed evidence that suggests small samples (e.g., less than 100) may lead to a biased estimate of χ^2, resulting in too

frequent rejections of the null hypothesis of χ^2 larger than zero. Other potential effects of a small sample are increased residuals and decreased normed fit indices. As a result, the relatively small sample used in the present experiment may have yielded an overly pessimistic view of the fit of the measurement and structural models to the data. The low GFIs provided by LISREL may be a function of the small sample. However, Bentler's CFI is also reported; this index avoids the underestiamtion of fit for small samples. The CFI values for the ability/performance models ranged from .894 to .951, which represent adequate fits (Bentler and Bonnett, 1980).

Simultaneous Equation Models

The current analytical procedure was designed to allow exploratory analysis of ability/performance relationships for CM, VM, and New CM after task-specific training and reversal performance. The separate specification searches yielded the best model of ability/performance relationships for each situation. Some discussion of similarities and differences across conditions was possible through the comparison of particular path coefficients along with general trends. Future analyses should include simultaneous equation models in which direct comparisons are made between and within conditions.

Limits of Exploratory Approach. Although structural equation modelling may be considered primarily a confirmatory approach, it is clearly useful as an exploratory approach as well (Hertzog, 1985, 1990). In the current experiment, general descriptive hypotheses were tested and subsequent model-fitting was performed based on the data (Alwin, 1988; Hertzog, 1985). The results of such a specification search have the potential to be sample-specific. Appropriate precautions must be taken in exploratory model-fitting to avoid capitalizing on chance.

First, only parameters that are theoretically meaningful (i.e., can be substantively supported) should be estimated based on inspection of the fit (or lack thereof) of the model. Second, only those paths that are statistically different from zero should be retained. Both procedures were followed in the present analysis. A third precaution is to assess the fit of the final model in an independent sample (i.e., cross-validation). Such a cross-validation was not conducted in the present experiment. The present results provide a baseline against which to compare future models of ability/performance relationships.

SUMMARY OF RESULTS AND CONCLUSIONS

The present experiment has provided an important first look at ability/performance relationships under CM and VM practice conditions on a visual search task. The results are consistent with the proposal that performance improvements in visual search are multifaceted and include general search strategies, optimal search strategies which may be stimulus-specific, and automatic response development. The pattern of results observed suggests that CM visual search performance improves as a function of all these components, but final-level CM performance is primarily a function of an automatic response to the target category. This conclusion is drawn from several convergent results. First, there was a decrease in RT with practice as well as a flattening of the comparison slope estimate for the CM task. Second, performance was severely disrupted (relative to final-level CM performance) when the roles of the CM targets and distractors were reversed. The disruption, more severe than that observed for the New CM condition, was a direct function of the number of distracting items in the display. Third, the ability/performance relationships for the practice data suggest that performance continues to improve and is related

to PS through the third session (1200 practice trials). Finally, additional influences of PS and SMA were necessary to predict transfer performance (relative to final-level CM performance); this suggests that the transfer conditions are functionally different than the trained CM condition.

The present experiment provides an existence proof of how normative analyses of mean RT data and structural models of the covariance patterns provide convergent information about the processes involved in performance. Combining the two approaches in a single experiment offered the unique opportunity to learn more about normative patterns of performance, individual differences and how they influence performance, and the stability of individual differences across practice. The results conform to previous suggestions that, in some situations, visual search performance is a function of differential mechanisms for CM and VM training. The present study also provides some evidence that learning within a given training procedure (CM or VM) is the result of multiple factors (i.e., general search strategies, optimal search strategies, and, for CM, target/distractor strengthening).

III. EXPERIMENTAL SERIES 2: LEARNING IN CONSISTENT SEARCH-DETECTION TASKS: TYPE OF SEARCH (MEMORY VS. VISUAL) DETERMINES TYPE OF LEARNING

Introduction

This section discusses an experiment which was conducted to understand and consolidate phenomena associated with learning and performance improvements in extended-practice search studies. Search paradigms have been a cornerstone of attention research for the past two decades (cf. Shiffrin, 1988) and have been important in gaining data to develop training principles (e.g., Fisk, Ackerman, and Schneider, 1987; Fisk, Hodge, Lee, and Rogers, 1990; Fisk, Rogers, Lee, Hodge, and Whaley, 1991). Our investigation centered on better understanding the relationship between the type of search processes used during training and what is learned during that training. An understanding of such search-related phenomena is important because they are closely related to attention issues in general and skill acquisition issues in particular.

To address the above issue, we investigated the effect of being trained with one procedure (e.g., memory search) on transfer to a different search procedure (e.g., visual search). Although similarities exist between memory scanning and visual search, these search processes seemingly involve different processing mechanisms (cf. Fisher, Duffy, Young, and Pollatsek, 1988; Flach, 1986; Hoffman, 1978, 1979). Pure visual search benefits most from the ability to differentiate targets from distractors (Duncan and Humphreys, 1989; Shiffrin and Czerwinski, 1988), whereas memory scanning is enhanced most by the ability to integrate the elements of a target set into a single equivalence class (Fisk and Schneider, 1983; Schneider and Fisk, 1984). These abilities are separable and appear to be dominated by different learning mechanisms (Schneider, 1985; Schneider

and Detweiler, 1987, 1988). A dissociation of the pattern of transfer effects in memory and visual search would support the distinction between memory and visual search.

Background: Search/Detection Procedure

In a typical search/detection experiment the subject is first presented with a memory set. The memory set contains the item(s) (target(s)) that the subject is to detect on a given trial. Following some time interval for study, the subject is presented with the display set. The display set contains the stimuli that the subject must compare to the memory set items. The display set may contain a memory set item (target), items not in the memory set (distractors), or both targets and distractors. The subject's task is to indicate the presence or absence of a memory set item, or to indicate the location of a memory set item within the display. Assuming that stimuli are visually presented (as in the current experiments), when memory-set size is one and display-set size is greater than one, experimenters are able to assess pure visual search. When memory-set size is greater than one and display-set size remains one, subjects' pure memory search ability can be tested. When both memory-set and display-set size are greater than one, the combined influence of memory and visual search is assessed (i.e., hybrid memory/visual search).

Performance Improvement in Search/Detection Tasks

Previous research has provided a solid knowledge base concerning general human performance characteristics within the realm of search/detection tasks (Fisk, Ackerman, and Schneider, 1987; Shiffrin, 1988; Shiffrin and Schneider, 1977). This empirical base has demonstrated the importance of the concept of consistency for performance improvement. For example, learning to shift gears in a car would be considerably more difficult if the gear-to-shifter location

changed every time you drove a car. Such inconsistency would necessitate continuous devotion of attention to remembering where the gears were; the task of shifting gears would show little learning and probably never become a skill. If a task is inconsistent, or varied, performance improvement is limited and skill will not develop.

Schneider and Shiffrin (1977; Shiffrin and Schneider, 1977) demonstrated differences in search performance as a function of whether training was consistent or varied. This has been referred to as consistently or variably "mapped" training. More precisely, in a consistent mapping (CM) situation the individual always deals with a stimulus, or a class of stimuli, in the same manner (i.e., the individual attends to, responds to, or utilizes information in the same manner across trials or task situations). CM training conditions result in dramatic performance improvements (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977), modifications in characteristics of event-related brain potentials (Kramer, Schneider, Fisk, and Donchin, 1986), and the eventual development of performance characteristics indicative of automatic processing. Varied mapping (VM) training situations are those in which the practice is inconsistent; that is, the response or degree of attention devoted to the stimulus changes from one stimulus exposure to another. VM training conditions result in little performance improvement other than that due to general familiarization with the task.

A Strength-Theoretic Perspective

Many skill development theories that envision performance and processing requirements changing as a function of practice are based on the modal view of a strength representation of knowledge (e.g., Anderson, 1982, 1983; Dumais, 1979; LaBerge and Samuels, 1974; MacKay, 1982; Schneider, 1985; Schneider and Detweiler, 1987; Shiffrin and

Czerwinski, 1988; but see Logan, 1988, for a non-strength theory). These theories all propose that some increase and/or decrease in "strength" is responsible for the performance improvement observed in tasks where substantive learning occurs (CM tasks within our paradigm).

The concept of strength varies among models, but is generally related to the importance or significance of a stimulus, set of stimuli, rule, or connection (e.g., between nodes). For example, MacKay's (1982) strength theory is based on repeated activation, priming, reinforcement, and resultant changes in the strength of connections between nodes. Production system models incorporate a concept of strength associated with production rules. The strength of a production is increased when the rule is invoked, and weakened when the application of the rule leads to error (Anderson, 1982, 1983). According to Neches, Langley, and Klahr (1987), "the strength (or weight) of a production is a parameter that is adjusted to indicate the system's current confidence in the correctness and/or usefulness of that rule" (p. 39). Finally, connection system models are strength-based in that they assume knowledge is stored in connection weights or strengths (Rumelhart and McClelland, 1987; Schneider and Detweiler, 1987).

Many investigations have provided evidence that supports the assumption that search performance is determined by the strength of the target relative to the strength of the distractor (e.g., Dumais, 1979; Prinz, 1979; Rogers, 1989). For the first trial, it is assumed that all stimuli have an equivalent, intermediate strength (Dumais, 1979; Shiffrin and Czerwinski, 1988; Shiffrin and Dumais, 1981). Note that the strength of the stimuli is intermediate and not zero because the stimuli are not completely novel but are simply untrained. For example, if words or letters are used as stimuli, they are familiar but

have not been previously trained to have a high strength level within the experimental context (Schneider and Fisk, 1984).

In general, each time a CM target appears in the display it is always attended and/or responded to. Therefore, the importance of a CM target increases and it is associated with a high "priority" tag (i.e., high strength level). After many trials, the high strength associated with CM targets will result in these items being processed without the need for serial search. Consistent distractors, on the other hand, will have a decreased strength level after practice because their appearance either results in a negative response (e.g., correct rejection) or no response at all. Therefore, CM distractors will have a very low priority. Finally, VM stimuli maintain an intermediate strength because on some trials they are targets and are attended to, but on other trials they serve as distractors and must be ignored. Conceptually, the strength of VM stimuli increases on some trials and decreases on other trials; therefore, even after many training trials these stimuli will still have an intermediate strength level. Studies examining the transfer and/or reversal of CM-trained targets and distractors reveal a pattern of results that supports strength-based theories of perceptual learning (e.g., Dumais 1979; Kristofferson, 1977; Rabbitt, Cumming, and Vyas, 1979; Rogers, 1989).

In addition to the development of attention-attracting strength, associative learning is a necessary component of performance improvement in search/detection tasks (Schneider and Detweiler, 1987). Our view of associative learning is not new; it has been precisely specified by other investigators (e.g., McClelland, Rumelhart, and Hinton, 1986; Schneider and Fisk, 1984; Shiffrin and Schneider, 1977). We assume memory to be a large collection of

interconnected nodes. Associative learning is reflected in the modification of the activation patterns between these nodes. Stimulus information, which is concurrently activated in short-term storage, will become associated if the co-activation consistently occurs across numerous learning trials. Once a set of information nodes becomes associated, a single representation can be extracted to represent the set. In our experiments, associative learning is assessed by the degree of categorization or unitization of the memory set. Such learning can facilitate search performance by allowing the comparison of a single, unitized memory set rather than forcing an item-by-item search through memory (Shiffrin, 1988).

Non-Strength Views of Performance Improvement

The above theoretical perspective of performance improvement assumes at least two mechanisms are responsible for the observed performance improvement in search/detection tasks: (1) target/distractor strengthening (i.e., targets attract attention and distractors repel it), and (2) memory factors (i.e., categorization or unitization develops for the memory set and possibly for the distractor set). Although other factors such as optimal feature search are important for improvement in visual search (Fisher, 1982, 1984; Fisher and Tanner, in press), it is important to note that a single mechanism is not proposed as the "cause" of performance improvement in search/detection studies.

However, other theoretical perspectives do propose single mechanisms. The most relevant for the present purposes is Logan's (1988; Logan and Klapp, 1991) automaticity-as-memory, or instance, theory of performance improvement. According to Logan, automaticity is associated with the development of efficient memory retrieval; that is, performance becomes automatic when it results from single-step, direct-access retrieval of past instances from memory.

Novice performance, he argues, is limited not by a lack of processing resources or lack of target/distractor strength differences, but by lack of knowledge. Learning, under this theoretical perspective, consists of acquiring specific solutions to specific problems. Logan allows for the possibility of generalization but offers no mechanism to account for such a process.

Logan's (1988) instance theory posits a discontinuity between novice and skilled performance. Initially, novices rely on a general algorithm that "generates" the solution to a particular problem or the performance of a skill. However, through repeated experience with specific problems, individuals learn specific solutions which they retrieve directly from memory. Subsequently, trainees can respond with a solution computed by use of the algorithm, or with one directly retrieved from memory. At some point, trainees abandon use of the algorithm and respond on every trial with solutions retrieved from memory. Thus, automatism (and performance improvement) represents the transition from algorithm-based to memory-based performance.

Extracting from Logan's (1988; July, 1991, personal communication) view of automaticity, instance theory would predict that if a subject is trained exclusively in memory search or trained exclusively in visual search, transfer to the other type of search would be poor. Instance theory assumes that the entire task situation leads to direct memory retrieval specific to the trained task. For example, instance theory would assume that the subject learns to retrieve specific representations of each display set in visual search.

Overview of the Experiment

In the present experiment, participants were trained in one of three CM search conditions: (1) pure memory search, (2) pure visual search, or (3) hybrid memory/visual search. Subsequent to the 6720 trials of practice, subjects transferred to a different search condition (or were not transferred and served in a "control" condition). For example, participants trained in pure memory search transferred to either pure visual search, hybrid memory/visual search, or continued to perform the pure memory search condition. Such transfer conditions were also created for the other two training conditions.

The present experiment is important because it will allow us to better evaluate the nature of learning during a given type of search/detection training. The pattern of transfer and the degree of that transfer across the various conditions should allow an evaluation of categorization and strengthening occurring (or not occurring) within each type of search. In addition, instance theory can be evaluated. If, in fact, a non-strength theory is sufficient to account for performance improvement (hence, strength is a useless construct), then no transfer should occur. If transfer is observed in at least one condition, we can suggest a limitation to instance theory.

Method

Subjects. Seventy-four volunteers completed the experiment. All subjects were undergraduates from Georgia Institute of Technology. Participants were tested for corrected or uncorrected far vision of at least 20/30 and near vision of at least 20/40. Subjects were given the option of receiving either course credit or $4.00 per hour for participation in the experiment.

Equipment. IBM personal computers were programmed with Psychological Software Tools' Microcomputer Experimental Language (Schneider, 1988) to present the appropriate stimuli, collect responses, and control the timing of the display presentations. Zenith monochrome monitors, controlled by a standard IBM CGS graphics adapter, were used to present the stimuli. The "1" and "2" numeric keypad keys were labeled "Y" and "N," respectively. All subjects were tested at individual subject stations which were monitored by an experimenter.

White noise was generated by a LaFayette (Model No. 15012) noise generator, amplified by a Denon (Model No. PMA-320) amplifier. The white noise was presented over two Sharp detachable two-way speakers located in the center of the testing room. The intensity of the noise at each workstation was approximately 72 dB (A).

Stimuli. Stimuli consisted of 12 semantically unrelated categories (except for flowers and vegetables which received a relatedness rating of 20-29 percent; Collen, Wickens, and Daniele, 1975). The categories used across the training and transfer phases of the experiment were: Building Parts, Clothing, Countries, Flowers, Four-Footed Animals, Human Body Parts, Musical Instruments, Natural Earth Formations, Occupations, Relatives, Vegetables, and Weapons. Target and distractor items were high associates of these categories (Battig and Montague, 1969; Howard, 1980). Each category contained eight words which were chosen because of the relatively equal confusability across words and categories. Within each condition, each subject received a unique order of category assignment as target and distractor sets for training and transfer. This counterbalancing was replicated across the between-subjects training conditions.

Design. During the training phase, the between-subject independent variable was Type of Training task, being either pure memory search, pure visual search, or hybrid memory/visual search. The within-subject variable was Sessions of Practice. The transfer phase contained nine unique, between-subject search conditions. Each previous Type of Training was factorially combined to produce the nine transfer search conditions. The nine conditions were:

1) Training on pure memory search with transfer to pure memory search (the M-M condition).

2) Training on pure memory search with transfer to pure visual search (the M-V condition).

3) Training on pure memory search with transfer to hybrid memory/visual search (the M-H condition).

4) Training on pure visual search with transfer to pure memory search (the V-M condition).

5) Training on pure visual search with transfer to pure visual search (the V-V condition).

6) Training on pure visual search with transfer to hybrid memory/visual search (the V-H condition).

7) Training on hybrid memory/visual search with transfer to pure memory search (the H-M condition).

8) Training on hybrid memory/visual search with transfer to pure visual search (the H-V condition).

9) Training on hybrid memory/visual search with transfer to hybrid memory/visual search (the H-H condition).

The within-subject independent variables during transfer were: (1) Type of categories being previous'y trained and new categories, and (2) Sessions of Transfer.

Procedure. An individual trial consisted of the following sequence of events. The participant was presented with the memory set (either one or three semantic category labels depending on the condition, see below), which he/she was allowed to study for up to 20 seconds. The memory set was displayed in the left-most, middle part of the display. Participants were instructed to press the spacebar to remove the memory set and initiate the remainder of the trial. One or three plus signs (again depending on the condition) were then presented for 0.5 seconds in the center of the screen; this allowed the participant to localize his/her gaze. Immediately following presentation of the fixation point, a visual mask was displayed to cover the memory set. The visual mask was followed by the display set (the response display), which consisted of one or three words (depending on the condition). The response display was presented for five seconds or until the subject responded - whichever occurred first. If one word was presented, it appeared in the center of the display. If three words were presented, they appeared in the middle of the display.

The participant's task was to determine whether an exemplar from the memory set was present in the response display. If the target was present, the participant pressed the key labeled "Y"; if the target was absent, the participant was instructed to press the "N" key. The probability of a target appearing was 0.50.

Each subject participated in seven one-hour training sessions and two one-hour transfer sessions. These nine sessions were conducted on nine consecutive days (including weekends). Within each session, 960 trials were presented in 20 blocks of 48 trials per block.

Participants received the following performance feedback. After each correct trial, the subject's RT was displayed in hundredths of a second. After each incorrect trial, an error tone was sounded and either the phrase "Error, target was 'target word'" (where "target word" was the exemplar from the memory set) or the phrase "Error, no target was in the display" appeared on the screen. At the end of each block of trials, the subject's average RT for correct trials and his/her average accuracy were presented. Subjects were instructed to maintain an accuracy rate of 93 percent. After each block of trials, the subject received one of the following three messages depending on his/her accuracy: (a) Your accuracy is below 92 percent. On the next block of trials please try to respond more carefully. (b) Your accuracy is above 96 percent. On the next block of trials please try to respond faster. (c) Your accuracy is fine.

The subjects were encouraged to take a break after each block of trials to rest their eyes. Prior to beginning a new session, each subject was privately given specific feedback by the experimenter regarding his/her performance on the previous day.

There were two phases of the experiment: training and transfer. In both phases, stimuli were consistently mapped as target and distractors.

Training Conditions. During training, subjects were assigned three of the 12 semantic categories as target categories and three as distractor categories. The training phase consisted of three conditions:

1) Pure Memory Search. In the pure memory search training condition, the subjects' memory set always consisted of the three target categories (that is, three category labels representing the

three target categories). The display set consisted of only one word. For target-present trials, the word was an exemplar from one of the target categories. For target-absent trials, the word in the display set was chosen from a category in the distractor set.

2) Pure Visual Search. In the pure visual search training condition, the subjects' memory set always consisted of only one of the three target categories (that is, one category label representing one of the potential target categories). All target categories (and words within the target categories) occurred an equal number of times across the training sessions. The display set consisted of three words. For target-present trials, one word was an exemplar from the category displayed in the memory set and the other two words were from the distractor categories. For target-absent trials, all three words were from the distractor categories.

3) Hybrid Memory/Visual Search. The hybrid memory/visual search condition is a combination of the other two training conditions. The subjects' memory set always consisted of the three target category labels and the display set always contained three words. For target-present trials, one of the words was an exemplar from the category displayed in the memory set; the other two words were from the distractor categories. For target-absent trials, all three words were from the distractor categories.

Transfer Conditions. For the two transfer sessions, the subjects transferred to (or continued to perform, for the baseline conditions) either pure memory search, pure

visual search, or hybrid memory/visual search. Hence, there were nine between-subject conditions tested during transfer (see the design section for details). In addition, for each transfer condition, a new CM condition was created from the six remaining categories that were not used during training (three categories were used as target categories and three were used as distractor categories). Thus, if a subject transferred to pure memory search, he/she would perform the memory search task using previously trained stimuli and, in a separate trial block, perform the memory search task using the new stimuli. The old and new stimulus condition was alternated between blocks of trials. The ordering of old categories first or new categories first was dependent on subject number (odd numbered subjects received old categories first). The transfer sessions each consisted of 20 blocks of 48 trials per block (960 trials per session). There were ten trial blocks (480 trials) using previously trained (old) stimuli and ten blocks using new stimuli.

Results

Data from the training and transfer sessions were analyzed separately. Below, we first present the RT and accuracy analyses for the training sessions, then present the analysis of the data from the transfer sessions. The transfer session data of three subjects were lost.

Training Data -- Reaction Time. The mean correct-trial RT scores for each search condition are presented as a function of sessions of practice in Figure 27. A Type of Training X Practice Sessions ANOVA with Subjects nested in Type of Training was conducted. (Note, an unequal subjects ANOVA was performed.) The analysis showed that the main effects of Type of Training and Sessions of Practice were significant;[12] $\underline{F}(2,71) = 57.61$ and $\underline{F}(6,426) = 508.67$,

[12] Unless otherwise indicated, α was set to .05.

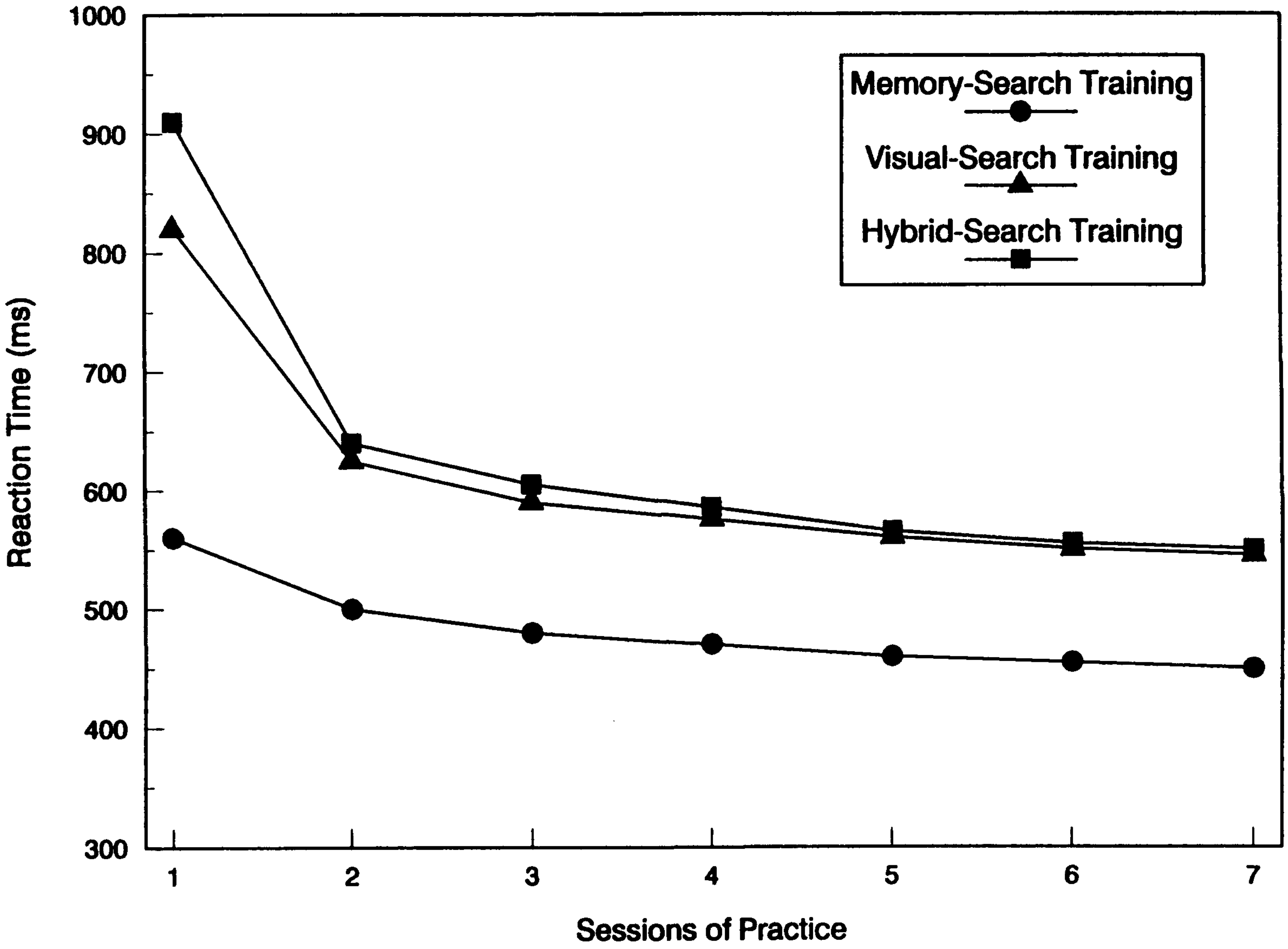


Figure 27. Mean Reaction Time as a Function of Practice Session and Type of Training

respectively. In addition, the interaction between Type of Training and Sessions of Practice was significant ($\underline{F}(12,426) = 47.30$). Figure 27 clearly shows the source of the main effect of Type of Training. Across all sessions, the pure memory search condition was faster than the pure visual search and the hybrid memory/visual search condition. Early in practice (Session 1) the pure visual search condition was significantly faster than the hybrid memory/visual search condition (827 ms vs. 917 ms, respectively); however, by the end of practice the difference was only 28 ms (523 ms vs. 551 ms, respectively). Both the pure visual and hybrid memory/visual search conditions improved more over practice than the pure memory search condition. Mean RT decreased by 304 ms over the course of the training phase for the pure visual search condition, and the hybrid memory/visual search condition showed a 365-ms decrease in mean RT. Because mean RT for the pure memory search condition in Session 1 was already very fast, that condition demonstrated a change in mean RT of only 106 ms. Hence, the interaction is due to the initially slower RTs of the pure visual and hybrid memory/visual search conditions relative to the pure memory search condition; although after 6700 practice trials there is still approximately a 100-ms difference between pure memory search and the other two search conditions.

<u>Training Data -- Accuracy</u>. The mean accuracy data are provided in Table 24 as a function of Type of Training and Sessions of Practice. Although the differences between training conditions and across practice are rather small, those main effects were statistically significant nonetheless. The ANOVA showed a significant main effect of Type of Training ($\underline{F}(2,71) = 5.51$) and a significant effect of Sessions of Practice ($\underline{F}(6,426) = 6.58$). The interaction did not reach significance ($\underline{F}(12,426) = 0.79$). The accuracy data clearly indicate that speed/accuracy trade-offs are not

Table 24. Accuracy for Training Phase: Section II. Accuracy Noted as Percent Correct.

	Type of Training		
Session	Pure Memory	Pure Visual	Hybrid Memory/Visual
1	94	93	93
2	94	92	92
3	94	92	92
4	93	92	92
5	93	92	91
6	93	92	91
7	93	92	92

present in such a way as to interfere with interpretation of the RT data. Because the pure memory search condition was both faster and more accurate than the other two conditions (although by a small amount), if anything we have underestimated the difference between the pure memory search condition and the other two conditions.

Transfer Data -- Reaction Time. Table 25 presents the mean RT data for all nine transfer conditions across the two sessions of practice for new and previously trained categories. A Transfer Condition X Type of Categories (Old or New) X Session ANOVA (with Subjects nested in Transfer Condition) was conducted. The main effects of Transfer Condition, Type of Categories, and Session were significant with $F(1,62) = 151.90$, $F(1,62) = 341.60$, and $F(8,62) = 15.23$, respectively. Two-way interactions between Transfer Condition and Type of Categories ($F(8,62) = 12.63$), Transfer Condition and Session ($F(8,62) = 5.40$), and Type of Categories and Session ($F(1,62) = 166.53$) were significant. Finally, the three-way interaction among Transfer Condition, Type of Categories, and Session reached significance ($F(8,62) = 6.02$).

The transfer data for the old, previously trained stimuli are presented in Figure 28. These data clearly indicate the effects of transfer as a function of previous search training. Figure 28 shows that for pure memory search training, transfer to either pure visual search or hybrid memory/visual search was disrupted. Memory search training provided some stimulus-specific benefit relative to searching for new stimuli; however, this benefit was short-lived. The difference between old and new stimuli diminished by the second transfer session for the previous pure memory search training group. Both visual search training and hybrid memory/visual search training resulted in extremely good transfer to the untrained search

Table 25. Reaction Time for Transfer Phase: Section II. Reaction Time Data (ms) for the Transfer Phase as a Function of Transfer Condition, Session, and Type of Categories (Old vs. New).

Transfer Condition			Transfer Session			
			1st		2nd	
Training	-->	Transfer	Old	New	Old	New
Memory	-->	Memory	447	527	443	485
Memory	-->	Visual	668	754	617	667
Memory	-->	Hybrid	682	819	628	681
Visual	-->	Memory	434	499	425	455
Visual	-->	Visual	538	750	529	670
Visual	-->	Hybrid	571	802	547	669
Hybrid	-->	Memory	473	555	485	530
Hybrid	-->	Visual	551	770	515	653
Hybrid	-->	Hybrid	575	889	574	736

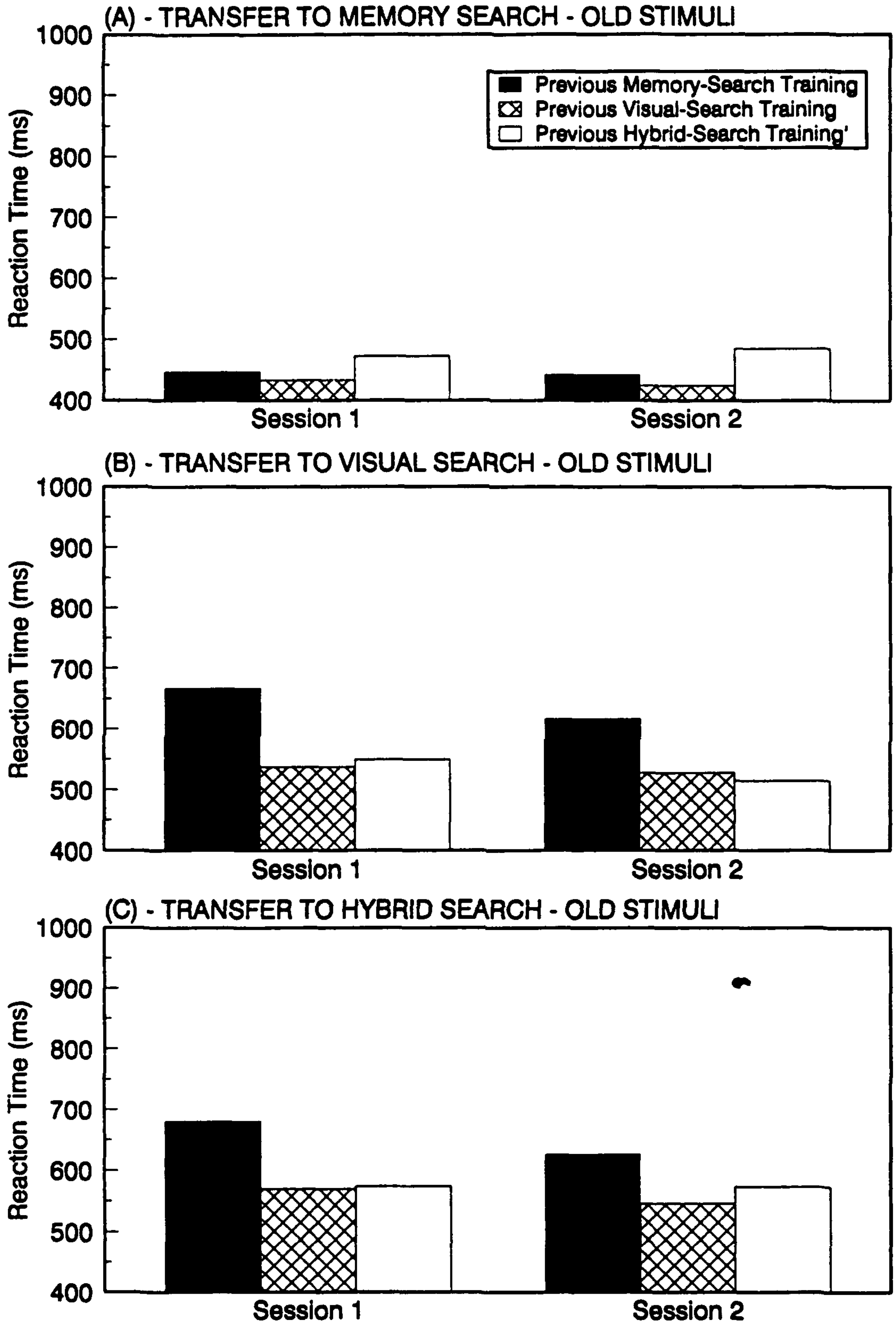


Figure 28. Mean Reaction Time as a Function of Transfer Type and Session (Old Stimuli, Transfer Session)

conditions. The benefit of the previous pure visual or hybrid memory/visual search training was still present in the second session of transfer (after almost 2000 practice trials).

Transfer Data -- Accuracy. The mean accuracy data are provided in Table 26 as a function of the nine transfer conditions across the two sessions of practice for new and previously trained categories. Once again, the small differences in accuracy resulted in statistically significant effects. A Transfer Condition X Type of Categories (Old or New) X Session ANOVA (with Subjects nested in Transfer Condition) was conducted. The main effects of Transfer Condition, Type of Categories, and Session were significant with $F(8,62) = 2.86$, $F(1,62) = 103.06$, and $F(1,62) = 9.07$, respectively. Two-way interactions between Transfer Condition and Type of Categories ($F(8,62) = 3.45$), Transfer Condition and Session ($F(8,62) = 1.94$), and Type of Categories and Session ($F(1,62) = 52.91$) were significant. Finally, the three-way interaction among Transfer Condition, Type of Categories, and Session reached significance ($F(8,62) = 2.60$).

Given the relatively small differences in accuracy, the relatively small F-ratios (except when Type of Categories is involved), and the significant three-way interaction, the accuracy differences should be interpreted with caution. Examination of Table 26 shows little reason for concern in terms of speed/accuracy trade-offs seriously altering the interpretation of the transfer RT data. In fact, for the most part, the conditions showing the slowest RT also exhibited the lowest accuracy and vice versa.

Discussion

Training Data. The training phase showed a striking difference between performance in conditions where load was

Table 26. Accuracy for Transfer Phase: Section II. Accuracy (Percent Correct) for the Transfer Phase as a Function of Transfer Condition, Session, and Type of Categories (Old vs.New).

Transfer Condition	Transfer Session 1st		2nd	
Training --> Transfer	Old	New	Old	New
Memory --> Memory	94	91	93	92
Memory --> Visual	88	87	92	92
Memory --> Hybrid	92	91	92	92
Visual --> Memory	94	91	92	92
Visual --> Visual	92	88	92	88
Visual --> Hybrid	92	86	91	88
Hybrid --> Memory	93	90	94	91
Hybrid --> Visual	94	90	92	91
Hybrid --> Hybrid	92	85	92	88

induced by memory set size versus display set size. These differences were especially striking early in practice. Within the first session of practice, performance in the pure memory search condition was 278 ms faster than performance in the pure visual search condition, and 367 ms faster than performance in the hybrid memory/visual search condition. Such findings suggest that, if transfer is not an issue (see below), task load should be increased via an increase in memory set size rather than an increase in visual set size (assuming no possibility for interaction effects with other tasks).

The major difference between memory and visual search (including hybrid memory/visual search) early in practice seems to be due to the speed with which memory search learning occurs. Within the first session of memory-search practice, 80 percent of the overall performance improvement for that condition had occurred.

This fast initial learning in pure memory search is not unexpected because it has been observed in previous studies. Fisk and Hodge (in press) have demonstrated this difference in improvement rates across search conditions. Fisk, Rogers, Giambra, and Rosenberg (1990) have shown that associative learning can occur quite rapidly. Salthouse and Somberg (1982, young subjects data) present CM memory-search data which showed that after one practice session (100 memory search trials per session), subjects improved very little over the next 50 practice sessions. Strayer and Kramer (1990) also demonstrated fast learning in memory search (Experiment 1) and showed that most of the improvement (in terms of reduction of memory load effects) can occur within the first few trials of practice in some situations (Experiment 2). In addition, Rogers (1991) has shown that a significant amount of learning occurs within the first practice session (also see Section II of this

report). Given that we collapsed over 20 blocks of practice, the first session of memory search data should be interpreted as representing fast learning.

The fast rate at which memory search improves can also account for why the visual search and hybrid memory/visual search conditions differ less than would be expected if comparison load, regardless of the source of that load, were used to predict early CM search performance. (Remember that in pure visual search, memory set size is one and display size is three giving a comparison load of three; however, in hybrid memory/visual search the comparison load is nine because both memory set size and display set size are three.) After one session of practice, pure visual search performance was 89 ms faster than hybrid memory/visual search; after two sessions of practice it was only 45 ms faster. If CM memory search shows relatively fast improvement then CM hybrid memory/visual search should quickly become dominated by the visual search load. This seems to be the case.

Transfer Data. The transfer data clearly show a dissociation between task structure used during training and subsequent ability to transfer to other types of search tasks. Subjects trained in pure visual search and hybrid memory/visual search were quite capable of transferring to any of the search conditions, including pure memory search. However, those individuals trained in pure memory search demonstrated limited transfer to either pure visual or hybrid memory/visual search. Clearly, although across the training conditions subjects saw the same stimuli and made consistent responses to those stimuli, the type of learning seems driven by the type of task. Although these points have been raised previously (Fisk and Rogers, 1991; Shiffrin, 1988), the empirical data have not been available

within search/detection tasks to directly address these hypotheses.

From a practical perspective, the present data suggest that it is important to provide pure visual search practice if individuals will be required to perform consistent tasks that sometimes require pure memory search and other times require pure visual search of the same material. Further, for the above mentioned situation, it may not be a requirement to also train under a memory search situation (even if one might ultimately be required to perform such a task) if visual search practice is given. Similarly, at least within the constraints imposed by the present experimental design, if CM hybrid memory/visual search may be required, pure CM visual search training could be a sufficient and more easily implemented training design.

Theoretical Implications. Shiffrin (1988) has nicely outlined the potential mechanisms responsible for automaticity in search/detection tasks. According to Shiffrin, at least two factors are responsible for performance improvement: (1) CM targets are "strengthened" to the point that automatic attention attraction develops (attention attraction is "weakened" for CM distractors such that they "repel" attention); and (2) stimulus set categorization develops such that all CM targets are unitized into an experimentally defined category and the category "label" is automatically extracted when a test item appears on the display (Shiffrin, 1988, p.758). When stimulus set categorization occurs, a test item need only be compared to this label which effectively reduces the memory set to one.

As Shiffrin (1988) noted, there is clear evidence that target/distractor strength differentiation occurs in visual search (which leads to the trained targets developing automatic attention responses). He further points to data

showing that consistent memory search leads to stimulus-set categorization. However, there are lingering questions that may be addressed by the present data. First, does differential target/distractor strengthening (attention attraction) develop for consistent stimuli in pure memory search? Second, does categorization (stimulus-set unitization) occur in pure visual search? Finally, what most limits performance in well-trained consistent hybrid memory/visual search?

If CM memory search practice produces both an automatic attention response (due to target strengthening) and stimulus-set categorization, then transfer performance to pure visual search should be as good as performance for those subjects trained in pure visual search. We found that transfer to visual search was poor when subjects received pure memory search training. Hence, we can conclude that automatic attention responses are not or, at best, weakly developed under pure memory search. The problem with differentiating between a position that does not assume target/distractor strength differentiation and one that assumes that target/distractor strength differentiation occurs but only weakly is that categorization may benefit visual search to a limited degree; this is a possibility that we cannot rule out. Therefore, we cannot differentiate between the development of a weak automatic attention response and the use of categorization to somewhat aid performance in visual search. However, the latter possibility seems of limited value in explaining performance improvement in visual search.

Does visual search lead to the development of an automatic attention response, stimulus-set unitization, or both? We could rule out the development of stimulus-set unitization if transfer from visual search to memory search had been poor. Unfortunately, from the perspective of

completely determining the underlying learning in visual search, transfer from visual search to memory search was very good. (In fact, subjects trained in visual search then transferred to memory search performed better--a nonsignificant 13 ms faster--at transfer than subjects trained in memory search.) Hence, we can conclude that either categorization is occurring during practice in CM visual search or a target "calling" strongly for attention, even in pure memory search, is sufficient to automatically identify it as a target. If the latter is true, we are faced with a finding that suggests we may not be able to isolate a mechanism that is both necessary and sufficient to perform memory search at an automatic level (attention strength may be sufficient but not necessary).

The transfer to hybrid memory/visual search sheds some light on the third question presented above. Because subjects trained in pure memory search did not show good transfer to hybrid memory/visual search and the subjects trained in pure visual search showed near perfect transfer, we can conclude the following. First, the visual search component seems to be the "limiting" factor in the CM hybrid search. Second, an automatic attention response seems necessary (and sufficient) to produce very good transfer to hybrid memory/visual search. Finally, because subjects trained in pure memory search showed poor transfer to the hybrid memory/visual search condition, we have further evidence that categorization plays a limited role (relative to an automatic attention response) in tasks requiring visual search.

The present data add to the ever increasing list of results that suggest different mechanisms are important for memory and visual search. The data also add to the list of situations that are not well-described by an instance-based (Logan, 1988) theory of automaticity. Instance theory, or

automaticity-as-memory, would predict the same transfer effects for pure memory search and pure visual search. The fact that subjects show perfect transfer from visual to memory search but demonstrate poor transfer from memory to visual search cannot be accounted for by instance theory. The present findings do not argue against an instance-based theory of automaticity; however, they do suggest further constraints to the applications of that theory.

Finally, we note that automaticity has been the center of some controversy (Cheng, 1985; Duncan, 1986; Durso, Cooke, Breen, and Schvaneveldt, 1987; Logan, 1988; Ryan, 1983; Schneider and Shiffrin, 1977, 1985; Shiffrin and Schneider, 1977, 1984). Logan and Klapp (1991) suggest that the controversies are due to "...the method of defining automaticity and not with the concept of automaticity itself" (p 191). We agree that the concept of automaticity is important. Further, we believe that the present data are important because they indicate that some recent definitions of automaticity may be incorrect.

IV. EXPERIMENTAL SERIES 3: THE EFFECTS OF INCONSISTENCY ON THE MAINTENANCE OF SKILL IN A SEMANTIC-CATEGORY SEARCH TASK

Introduction

The purpose of the present investigation was to examine the effect of changes in the degree of task consistency on performance in a well-learned task. Within a semantic category visual search paradigm, highly trained performers had to adjust to different degrees of task inconsistency. In addition, the effects of varying degrees of task inconsistency on the maintenance of skilled search was investigated.

Understanding how people become skilled is critical information for the design of training programs. A fruitful approach to such an understanding has involved the study of skill development in visual search tasks (e.g., Ackerman, 1988; Fisk, Ackerman, and Schneider, 1987; Fisk and Rogers, 1988; Myers and Fisk, 1987; Shiffrin and Dumais, 1981). In visual search tasks, subjects typically detect target stimuli presented among irrelevant nontargets (distractors). A visual search task was chosen for the proposed experiment because it is a well-studied paradigm (Shiffrin, 1988). Results from research in this area have been successfully applied to the training of cognitively demanding skills (Eggemeier, Fisk, Robbins, and Lawless, 1988; Fisk and Eggemeier, 1988; Halff, Hollan, and Hutchins, 1986). Also, as described previously in this report, results from visual search studies have also been useful for formulating general theories of visual information processing (Shiffrin, 1988).

Researchers have recommended that training focus on the consistencies operating in the task environment (Eggemeier, Fisk, Robbins, and Lawless, 1988; Fisk and Eggemeier, 1988; Schneider, 1985). The logic for such recommendations is

straightforward: training on consistent task components leads to automatic processing of those components and, once learned, those automatically processed components will transfer to performance in the operational environment. In reality, however, even a well-designed training program cannot anticipate all possible real-world situations. Specifically, the consistent components encountered during training may, at times, become less consistent in the operational environment. Therefore, it is important to know what happens to skilled performance when the consistent aspects of a task change and become inconsistent.

Previous research has demonstrated performance disruption for the permanent reversal of targets and distractors (see Section II of this report for a discussion of reversal procedures and reversal effects). One unexamined situation is the temporary reversal of targets and distractors on the maintenance of skilled performance. We can address this issue by first providing subjects with extensive training on a CM task, manipulating the degree of consistency, then returning the subjects to the original CM training procedure.

Degree of search consistency has been defined as the number of trials within a block that an item appears as a target relative to the number of times an item appears as a distractor (Schneider and Fisk, 1982a). Schneider and Fisk trained subjects to detect stimuli (letters) that were either 100, 67, 50, 33-percent consistent during training. Those researchers demonstrated that the development of skilled performance in a detection task was a function of both the degree of consistency and the amount of practice. In essence, the Schneider and Fisk degree-of-consistency manipulation can be thought of as training subjects with varying levels of reversal trials. However, the effect of

varying degrees of consistency on established skills has not been investigated.

The goal of the present experiment was to examine how different degrees of consistency interfere with performance once a skill has been developed, and what long-term effects such inconsistency has on the maintenance of the skill. Specifically, the experiment examined the effects of degrees of consistency on performance and maintenance of performance in a well-learned semantic-category visual search task. In the present experiment, subjects were first given CM training to develop skill in visual search. Subjects then transferred to either 100, 67, 50, or 33-percent consistent search. Following practice in the degree of consistency phase, subjects returned to 100-percent consistent search conditions. The return to consistent search allowed us to examine the effects of the degree of consistency on the maintenance of skilled performance.

The present experiment will also allow us to examine the importance of various possible mechanisms of skill development (e.g., context, optimal search strategies, and automatic processing). In addition, we will determine whether a small or large degree of inconsistency is needed to disrupt skilled search. Finally, we will be able to address whether the inability to use an automatic process (or the need to inhibit an automatic process) disrupts other automatic processes on non-disrupted tasks.

Method

Subjects. Thirty-two students from a southeastern university (24 males and eight females) participated in this experiment. Participants ranged from 18- to 34-years-old with an average age of 20.53 years. Students received course credit or $4.00 per session for their time. All

participants had visual acuity of at least 20/30 (far vision) and at least 20/40 (near vision).

Stimuli. The stimuli for the present experiment were eleven semantically unrelated (Collen, Wickens, and Daniele, 1975) categories of Furniture, Vehicles, Trees, Clothing, Weapons, Earth Formations, Units of Time, Occupations, Vegetables, Relatives, and Alcohol. Six high associates from each category (Battig and Montague, 1969), four to seven letters long, were chosen as exemplars.

A calibration study was conducted to find a set of relatively equally confusable stimuli. Based on that study, the categories Furniture, Vehicles, and Trees were chosen as target categories.[13] Target categories were randomly assigned to search conditions. The assignment of distractor categories to phase of the experiment was counterbalanced using a partial Latin Square.

Apparatus. Microcomputers were programmed to control the timing of the displays, present the stimuli, and collect responses. All computer programs were developed using Psychological Software Tools' Micro Experimental Laboratory software (Schneider, 1988). The data were collected using three EPSON Equity I+ microcomputers with EPSON MBM 2095-5 green monochrome monitors, and one AGI 1800 AT-compatible microcomputer with a Goldstar 1430 VGA monitor (white text on a black background).

The EPSON Q-203A keyboard was altered by exchanging the "7," "4," and "1" numeric keypad keys with the "T," "M," and "B" keys, respectively. The "7," "4," and "1" numeric

[13] A calibration experiment suggested that similar reaction times to the categories Furniture, Vehicles, and Trees (750, 766, and 747 ms, respectively) would occur early in practice when paired with the distractor categories Clothing, Weapons, Earth Formations, Units of Time, Occupations, Vegetables, Relatives, and Alcohol.

keypad keys were marked with the paper labels "T," "M," and "B," respectively, on the AGI 1800 keyboard. For all experimental sessions, pink noise played at approximately 57db(A) to attenuate background noise. All subjects were tested at individual workstations which were monitored by a laboratory assistant.

Procedure. During the first session, subjects were given instructions and an orientation to the task which consisted of 150 CM trials. These orientation trials allowed the subjects to become familiar with the experimental protocol and also served to stabilize the error rates. The categories used for the orientation trials (i.e., Colors and Birds) were not used in the remainder of the experiment.

For each trial, the memory set contained one category label (memory-set size of one) and the display contained three words (display-set size of three). An individual trial consisted of the following sequence of events. The subject was presented with one category label as the memory set item which he/she was allowed to study for a maximum of 20 seconds. Subjects were instructed to press the space bar to initiate the trial. Three plus signs were then presented for 0.5 seconds in the center of the screen to allow the subject to localize his/her gaze. A target (i.e., an exemplar from the memory-set category) was present on every trial. The display set consisted of three words presented in a column. The subject's task was to indicate the location of the target (i.e., top, middle, or bottom) by pressing the corresponding key (labeled "T," "M," or "B").

Feedback. Participants received performance feedback at the end of each trial, the end of each block of fifty trials, and at the start of each session. After correct trials, RT was displayed in milliseconds. After incorrect trials, an error tone sounded, followed by a display of the

correct response. At the end of each block (50 trials), the subject's average RT and accuracy for that block were presented.

At the end of every block of trials, subjects received feedback concerning their accuracy. If a subject's mean accuracy (for that block) fell below 92 percent, a message encouraging him/her to respond more carefully on the next occurrence of that type of block was displayed. If the subject's mean accuracy exceeded 96 percent, a message encouraging him/her to respond faster on the next occurrence of that type of block was displayed. If the subject received either message, he/she was also reminded at the start of the next block of that type of trials with the message, "Remember to respond [carefully or faster] on this group of trials."

Subjects were encouraged to respond as fast as possible, while keeping their accuracy within the range described above. Before each session, participants privately received feedback on their previous day's performance. Each subject was shown his/her RT, presented in a graph, and encouraged to improve his/her performance from session to session.

<u>Experimental Sequence</u>

The experiment was conducted in three phases: Training, Degree of Consistency, and Retraining. Phase 1, the Training Phase, provided pure CM search practice. Previous research clearly indicates that such training allows subjects to become skilled at the semantic category search task (e.g., Rogers, 1991; and Section II of this document). The Training Phase was followed by Phase 2, the Degree of Consistency Phase. This phase examined the effect of introducing various levels of inconsistency on skilled visual search performance. Phase 3, Retraining, reinstated

the pure CM training procedure. The third phase of the experiment is referred to as the Retraining Phase. The Retraining phase was designed to examine the effect of the previous degree of consistency manipulation on the maintenance of skilled performance. The dependent measures in all three phases were RT and accuracy. A summary of the experimental sequence and the conditions within each phase are presented in Table 27.

Training Phase, Experimental Design

In the training phase all 32 subjects received training on two CM categories. Two categories were consistently mapped as target categories. One of the target categories is referred to as the Continuously-Consistent category because it was consistently mapped as a target throughout the entire experiment (all three phases). This category was included as a within-subject control for Phases 2 and 3. For the other target category, the Adjusted-Consistent category, consistency was adjusted, or changed, as a function of the phase of the experiment and the Degree-Group (see below) to which the subject was assigned. For the Adjusted-Consistent category, consistency was 100 percent during the Training Phase of the experiment, adjusted as a function of Degree-Group in the Phase 2, and 100 percent during the final phase. (See Table 28 for a summary of the Training Category consistency as a function of experiment phase and Degree-Group.) From the three possible target categories (Furniture, Vehicles, and Trees) for each subject, one category was randomly assigned as the Continuously-Consistent category and another as the Adjusted-Consistent category. The remaining target category was used as a new CM target category in Phase 3 (see below for discussion of Phase 3).

Two categories (from the set of possible distractor categories) made up the distractor set. The assignment of

Table 27. Training and Transfer Condition Summary

Training Phase		Degree of Consistency Phase		Retraining Phase	
Degree-Group 100					
Training Category	# of Trials[a]	Training Category	# of Trials	Training Category	# of Trials
C-C[b]	500	C-C	200	C-C	350
A-C[c]	500	A-C	200	A-C	350
		VM[d]	600	New CM	350
Degree-Group 67					
Condition	# of Trials	Condition	# of Trials	Condition	# of Trials
C-C	500	C-C	200	C-C	350
A-C	500	A-C	200	A-C	350
		Reversal[e]	100	New CM	350
		VM	500		
Degree-Group 50					
Condition	# of Trials	Condition	# of Trials	Condition	# of Trials
C-C	500	C-C	200	C-C	350
A-C	500	A-C	200	A-C	350
		Reversal	200	New CM	350
		VM	400		
Degree-Group 33					
Condition	# of Trials	Condition	# of Trials	Condition	# of Trials
C-C	500	C-C	200	C-C	350
A-C	500	A-C	200	A-C	350
		Reversal	400	New CM	350
		VM	200		

[a] Number of trials indicates number of trials per session.
[b] C-C represents the Continuously-Consistent Category.
[c] A-C represents the Adjusted-Consistent Category.
[d] VM represents Supplemental-VM.
[e] Reversal refers to those trials for which a word from the Adjusted-Consistent category serves as a distractor item.

Table 28. Degree of Consistency for Each Condition: Section III. Percent Consistency for Each Training Category as a Function of Degree-Group for Each Phase of the Experiment

	Phase of the Experiment		
Degree-Group	Training	Degree of Consistency	Retraining
Degree-Group 100			
C-C[a]	100	100	100
A-C[b]	100	100	100
VM	--	23	--
New CM	--	--	100
Degree-Group 67			
C-C	100	100	100
A-C	100	67	100
VM	--	25	--
New CM	--	--	100
Degree-Group 50			
C-C	100	100	100
A-C	100	50	100
VM	--	27	--
New CM	--	--	100
Degree-Group 33			
C-C	100	100	100
A-C	100	33	100
VM	--	33	--
New CM	--	--	100

[a] C-C represents the Continuously-Consistent Category.
[b] A-C represents the Adjusted-Consistent Category.

distractor categories to a given phase of the experiment was partially counterbalanced using a Latin Square. On each trial, the display set consisted of three words: a target word and two distractor words (one distractor chosen at random from each distractor category).

There were six sessions in the Training Phase. Each session consisted of 1000 trials (500 trials per target category) for a total of 6000 training trials. In preparation for the Degree of Consistency Phase of the experiment, the 32 subjects were randomly assigned to one of the four between-subjects groups (Degree-Group) with the restriction that each group have the same proportion of males and females (six males and two females per group). Each group received the same training during the Training Phase. The Degree-Groups were: Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33.

To summarize the design, the within-subject independent variables were Training Category (Continuously-Consistent and Adjusted-Consistent) and Practice (six sessions of practice). The between-subject independent variable was Degree-Group (Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33). However, degree of consistency was not manipulated in the first phase; hence, Training Category and Degree-Group should not have an effect on performance in this phase unless the random assignment of subjects and/or categories produces a bias. (Should a bias occur, the within-subject controls will still afford meaningful interpretation of the data.) The Training Phase was provided to develop an "automatic process" to both Training Categories in all Degree-Groups prior to Phase 2 manipulations.

Degree of Consistency Phase, Experimental Design

In this phase, the between-subjects, degree-of-consistency manipulation was introduced. There were four training sessions, each consisting of 20 blocks of 50 trials per block (1000 trials per session). As indicated above, the Continuously-Consistent category remained a consistent target category and the Adjusted-Consistent category underwent a degree of consistency manipulation. A Supplemental-VM condition (four categories) was also introduced.

To summarize the design of Phase 2, the within-subject independent variables were Training Category (Continuously-Consistent, Adjusted-Consistent, and Supplemental-VM) and Practice (four practice sessions). (For Degree-Group 67, Degree-Group 50, and Degree-Group 33 there were also "reversal" trials; these reversal trials are explained below.) The between-subject independent variable was Degree-Group (Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33). The Training Categories were presented in randomly ordered blocks (50 trials per block). The Training Category levels are outlined next.

Continuously-Consistent Condition (Within-Subject Control Condition). For all groups, one of the two target categories (Continuously-Consistent) used in Phase 1 remained consistently mapped as a target category. The Continuously-Consistent condition served as a within-subject control. There were four blocks of the Continuously-Consistent condition (a total of 200 trials) in each session. For a given Continuously-Consistent condition trial, distractors were chosen from two of the four Supplemental-VM categories.

Adjusted-Consistent Condition (Degree of Consistency Manipulation). In Phase 2, the Adjusted-Consistent category

underwent a degree-of-consistency manipulation. The degree of consistency was manipulated between subjects and was either 100, 67, 50, or 33-percent consistent with the degree of consistency of the Adjusted-Consistent condition determined by the Degree-Group to which a subject was assigned.

Degree of consistency of the Adjusted-Consistent category was determined by varying the ratio of times the category words appeared as targets versus distractors. The frequency with which the category appeared as a target was equal, whereas the frequency of its appearance as a distractor was varied. Within each of the four groups, there were 200 target trials per session for the Adjusted-Consistent category. The number of times the Adjusted-Consistent category served as a distractor was varied across Degree-Groups. Let (t:d) represent the number of times the Adjusted-Consistent category appeared as a target (t) versus a distractor (d). For Degree-Group 100, the target/distractor ratio was 200:0 per session. For Degree-Group 67, Degree-Group 50, and Degree-Group 33 the target/distractor ratio was 200:100, 200:200, and 200:400, respectively. The trials for which a word from the Adjusted-Consistent category occurred as a distractor are referred to as <u>Reversal Trials</u> (these trials are actually "half-reversal" trials, see Dumais, 1979; Rogers, 1989).

Within a block of Adjusted-Consistent trials, the Adjusted-Consistent category served either as a target, or a distractor on a particular trial (as defined above). When the Adjusted-Consistent category served as a target two distractor words were chosen at random, one each from two of the remaining Supplemental-VM categories. When a word from the Adjusted-Consistent category appeared as a distractor, the memory set contained a Supplemental-VM category. The

other distractor for that trial was randomly chosen from one of the three remaining Supplemental-VM categories.

Degree of consistency was defined across a session; however, the degree of consistency in each block either matched or closely approximated the overall degree of consistency within the session. For each block of trials for Degree-Group 100, the Adjusted-Consistent category served as a target in 50 (100 percent) of the trials. For Degree-Group 50, the Adjusted-Consistent category served as a target 25 times and as a distractor 25 times per block.

For a given block of trials, the overall degree of consistency could only be approximated for Degree-Group 67 and Degree-Group 33. Therefore, in each block of Adjusted-Consistent trials the degree of consistency category served as a target in at least 30 trials and as a distractor in at least 15 trials for Degree-Group 67. For Degree-Group 33, the Adjusted-Consistent category served as a target in at least 15 trials and as a distractor for at least 30. For the remaining five trials in each block, whether the Adjusted-Consistent category served as target or as a distractor was determined randomly, with the constraint that the Adjusted-Consistent category serve as a target 200 total times per session, and as a distractor either 100 or 400 times per session, for Degree-Group 67 or Degree-Group 33, respectively.

To create the appropriate degree of consistency (while keeping the number of trials per block equal across conditions), it was necessary to vary the number of Adjusted-Consistent trial blocks across Degree-Group. Subjects participating in Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33 received four, six, eight, and 12 blocks of Adjusted-Consistent trials, respectively. Because all Degree-Groups received an equal number of total trials per session, each Degree-Group also

received a different number of Supplemental-VM trial blocks (see below).

Supplemental-Varied-Mapping Search Condition

In each session of the Degree of Consistency Phase, subjects performed either 12, ten, eight, or four blocks of Supplemental-VM search (600, 500, 400, or 200 total Supplemental-VM trials per session, for Degree-Group 100, Degree-Group 67, Degree-Group 50, or Degree-Group 33, respectively).

In the blocks of Supplemental-VM trials, four categories that were not used in the Training Phase served as both targets and distractors. When a word from any one Supplemental-VM category was the target word, two distractor words were chosen at random, one each from two of the remaining Supplemental-VM categories. Within the Supplemental-VM condition, each of the Supplemental-VM categories served as a target category 150, 125, 100, or 50 times and as a distractor category 300, 250, 200, or 100 times for Degree-Group 100, Degree-Group 67, Degree-Group 50, or Degree-Group 33, respectively. However, because the Supplemental-VM items were used as distractors for the Continuously-Consistent blocks and the Adjusted-Consistent blocks (as well as targets for some trials within the Adjusted-Consistent blocks), the actual degree of consistency of the Supplemental-VM items was 0.23, 0.25, 0.27, and 0.33 for Degree-Group 100, Degree-Group 67, Degree-Group 50, or Degree-Group 33, respectively. Given that the degree of consistency for the Supplemental-VM was higher than that normally used for VM conditions, the Supplemental-VM condition was not a VM condition as typically designed in research examining CM/VM training effects. The Supplemental-VM stimuli were included to facilitate the degree of consistency manipulation.

Retraining Phase, Experimental Design

In Phase 3 the pure CM procedure was reinstated for all target categories. Degree of consistency (the between-subjects variable from the second phase) was not manipulated in the Retraining Phase; hence, all four groups received CM training on all target categories.

In Phase 3, subjects completed four sessions of training (1050 trials per session). Each session consisted of 21 blocks of 50 trials per block. The Training Category variable now consisted of: (1) the Continuously-Consistent category, (2) the Adjusted-Consistent category, and (3) a New CM category (this category was not used in the first two phases of the experiment). There were seven blocks of each search condition per session with presentation order randomly determined. Two new categories made up the distractor set. For each trial, the display set consisted of a target word and two distractor words (one distractor chosen at random from each distractor category).

To summarize the design, the within-subject independent variables were Training Category (Continuously-Consistent, Adjusted-Consistent, and New CM) and Practice (four sessions of retraining practice). The between-subject independent variable was Degree-Group (Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33). However, degree of consistency was not manipulated in Phase 3.

Results

Training Phase

Reaction Time. The mean, correct-trial RTs for Sessions 1 through 6 are presented in Table 29 for each degree of consistency Degree-Group. To examine improvement across training, a Degree-Group X Training Category X Practice (4 X 2 X 6) ANOVA was performed on the correct-

Table 29. Mean Reaction Time (ms) for Each Training Category by Degree-Group, in Sessions 1 Through 6 (Training Phase)

Training Category	Session					
Continuously-Consistent	1	2	3	4	5	6
Degree-Group 100	686	603	587	562	547	523
Degree-Group 67	644	585	553	547	529	515
Degree-Group 50	666	588	572	557	536	520
Degree-Group 33	679	599	588	565	532	519

	Session					
Adjusted-Consistent	1	2	3	4	5	6
Degree-Group 100	637	564	550	536	524	502
Degree-Group 67	636	572	542	533	510	504
Degree-Group 50	681	594	564	542	526	508
Degree-Group 33	665	587	569	541	509	500

trial RTs. The main effect of Degree-Group was not significant ($F(3,28) = 0.15$).[14] This nonsignificant effect is expected if subject selection did not introduce biases across the between-subject groups. RTs improved over the six sessions of practice ($F(5,140) = 205.86$). There were no significant interactions among any combination of Training Category, Practice, or Degree-Group.

At the end of the Training Phase, the Adjusted-Consistent condition was faster than the Continuously-Consistent condition ($F(1,28) = 5.64$). The target categories used in the Training Phase (and subsequently in the other phases of the experiment) were selected based on pilot data that showed them to yield roughly equal latencies in a Supplemental-VM design. However, in the present experiment the Furniture category unexpectedly yielded faster RTs compared to the Vehicles and Trees categories. Random assignment of categories to Training Category resulted in more subjects receiving the Furniture category as the Adjusted-Consistent category. Hence, this unfortunate category assignment seems responsible for the difference between Training Category conditions. The Training Category effect will cause the underestimation of the performance disruption caused by the degree of consistency manipulations; however, the within-group controls built into the experiment will allow an unbiased estimate of the degree-of-consistency manipulation in Phases 2 and 3.

Training Phase, Accuracy. A Degree-Group X Training Category X Practice (4 X 2 X 6) ANOVA was conducted on the accuracy data from the Training Phase. Analysis of the accuracy data showed a significant Degree-Group X Training Category X Practice interaction ($F(15,140) = 1.89$). No

[14] Unless otherwise indicated, α is 0.05. In addition, all analyses involving repeated measures are evaluated using the Greenhouse-Geiser correction.

other sources of accuracy variance reached significance. Mean accuracy for each Degree-Group and Training Category are presented in Table 30 as a function of practice session. During training, accuracy fluctuated between 92 and 95 percent. At the end of training, there were no significant differences among the Degree-Groups ($F(3,28) = 1.19$) or Training Categories, ($F(1,28) = 1.70$), and the interaction between Degree-Group and Training Category, failed to reach significance ($F(3,28) = 0.91$).

Training Phase, Summary. In Phase 1 of the experiment, 11 Degree-Groups improved--as generally would be expected with CM training (performance was well-described by a power function, $RT_i = 280 + bp_i^{-.2}$, where b is the initial RT and p_i is the given point in practice where RT is estimated). At the end of training, the Degree-Groups did not differ. The Adjusted-Consistent condition was faster than the Continuously-Consistent condition. Accuracy did fluctuate during training, but within instructed limits. At the end of training, there were no accuracy differences among the Degree-Groups or between the Training Categories.

Table 30. Mean Percent Accuracy for Each Training Category by Degree-Group, in Sessions 1 and 6 (Training Phase)

Training Category	Session					
Continuously Consistent	1	2	3	4	5	6
Degree-Group 100	93	93	93	93	94	93
Degree-Group 67	93	93	93	93	93	93
Degree-Group 50	94	94	94	94	94	94
Degree-Group 33	93	93	93	93	92	93

	Session					
Adjusted Consistent	1	2	3	4	5	6
Degree-Group 100	94	94	93	93	94	93
Degree-Group 67	93	93	94	94	93	93
Degree-Group 50	94	94	95	94	94	94
Degree-Group 33	94	94	93	94	93	93

Degree of Consistency Phase

Reaction Time, Disruption at Initial Session. Mean correct-trial RTs for the Continuously-Consistent condition and Adjusted-Consistent condition are presented in Table 31 as a function of Degree-Group. The data in Table 31 for Adjusted-Consistent condition are for target trials only (reversal trials will be discussed in another section). For comparison purposes, the average RTs for each condition at the end of the previous Training Phase are also shown in Table 31. Mean difference scores are presented to illustrate the amount of change experienced when subjects transferred to the Degree of Consistency Phase.

Clearly, transfer to this second phase of the experiment slowed the subjects' RT. RT for the Continuously-Consistent condition increased between 70 and 90 ms but was not related to Degree-Group. Slowing in the Adjusted-Consistent condition was directly related to Degree-Group, ranging from 90 ms to 224 ms. For all Degree-Groups, performance in the Adjusted-Consistent condition was worse than performance in the corresponding Continuously-Consistent condition (except, of course, for Degree-Group 100--the group that maintained 100 percent consistency in both Training Categories). A Degree-Group X Training Category X Phase (end of Training phase versus beginning of Degree of Consistency phase) ANOVA revealed a significant Degree-Group X Training Category X Phase interaction, ($F(3,28) = 3.57$). This analysis confirms an inspection of Table 31 that disruption of performance in the Adjusted-Consistent condition is related to Degree-Group participation.

RTs for the Continuously-Consistent and Adjusted-Consistent conditions for each Degree-Group were compared directly. Those data, and the corresponding difference scores, are presented in Table 32. Recall that the

Table 31. Mean Reaction Time (ms) and Difference Scores for the Continuously-Consistent and Adjusted-Consistent Categories by Degree-Group, in Sessions 6 and 7 (Training Phase to Degree of Consistency Phase Transfer)

	Session		
Continuously-Consistent	6	7	Difference
Degree-Group 100	523	608	-85
Degree-Group 67	515	609	-94
Degree-Group 50	520	591	-71
Degree-Group 33	519	609	-90

	Session		
Adjusted-Consistent	6	7	Difference
Degree-Group 100	502	592	-90
Degree-Group 67	504	671	-167
Degree-Group 50	508	654	-146
Degree-Group 33	500	724	-224

Table 32. Mean Reaction Time, Percent Accuracy, and Difference Between the Adjusted and Continuously-Consistent Categories by Degree-Group, in the First Session of the Degree of Consistency Phase (Session 7)

	Reaction Time		
	Continuously-Consistent	Adjusted-Consistent	Difference
Degree-Group 100	608	592	16
Degree-Group 67	609	671	-62
Degree-Group 50	591	654	-63
Degree-Group 33	609	724	-115

	Accuracy		
	Continuously-Consistent	Adjusted-Consistent	Difference
Degree-Group 100	93	93	0
Degree-Group 67	93	95	-2
Degree-Group 50	94	95	-1
Degree-Group 33	93	93	0

comparison of RTs between the Continuously-Consistent and the Adjusted-Consistent conditions is biased against finding disruption in the Adjusted-Consistent condition because that condition was significantly faster than the Continuously-Consistent condition after training in Phase 1. This caution notwithstanding, the difference between the Continuously-Consistent and the Adjusted-Consistent conditions was statistically significant ($\underline{F}(1,28) = 13.04$). Importantly, even with this biased measure, performance degradation in the Adjusted-Consistent condition is related to degree of consistency. This finding is supported by the significant interaction of Degree-Group X Training Category ($\underline{F}(3,28) = 3.02$).

As another measure of the effect of varying degree of consistency on search performance, we calculated the percentage of disruption for each Adjusted-Consistent condition <u>relative</u> to each group's performance in the corresponding Continuously-Consistent condition using the following formula:

$$RD = (1 - ((BCS_{rt} - ACS_{rt})/(BDC_{rt} - ADC_{rt}))) * 100,$$

where RD is relative disruption, BCS_{rt} is RT for the Continuously-Consistent condition at the end of training, ACS_{rt} is RT for the Continuously-Consistent condition at Session 1 of the Degree of Consistency phase, BDC_{rt} is RT for the Adjusted-Consistent condition at the end of training, and ADC_{rt} is RT for the Adjusted-Consistent condition at Session 1 of the Degree of Consistency Phase.

The relative disruption scores are in line with the predicted degree of consistency effects. The averaged relative disruption was 5.6, 43.7 51.4, and 59.8 percent, respectively, for Degree-Group 100, Degree-Group 67, Degree-Group 50 and Degree-Group 33. Even with this measure, we

find that target detection of the well-trained stimuli is disrupted as a direct function of degree of consistency.

Degree of Consistency Phase, Accuracy During Initial Session. Percentage correct detection, for the Continuously-Consistent condition and the Adjusted-Consistent condition, for each Degree-Group, are also presented in Table 32. The Degree-Group X Training Category interaction reached significance ($F(3,28) = 4.07$). Subjects made fewest errors in Degree-Group 50 and Degree-Group 67. These differences are small (one percent change in accuracy in a session corresponds to a difference of two errors) despite the statistical significance. Importantly, the accuracy data do not suggest that speed-accuracy trade-off effects preclude our discussion of the RT data.

Degree of Consistency Reaction Time, All Sessions. Mean, correct-trial RTs for all Degree of Consistency sessions are presented in Table 33. An ANOVA examining Degree-Group X Training Category X Practice (4 X 2 X 4) was conducted. RT performance improved on average across the four sessions as indicated by a significant main effect of Practice ($F(3,84) = 21.43$). The interaction between Training Category and Degree-Group was also significant ($F(3,28) = 3.75$). Practice did not interact with any other variables. The Degree-Group X Training Category X Practice interaction ($F(9,84) = 1.24$), the Degree-Group X Practice interaction ($F(9,84) = 1.32$), and the Training Category X Practice interaction ($F(3,84) = 1.28$) all failed to reach significance. Hence, we conclude that the pattern of effects found in the initial session of the Degree of Consistency Phase remained relatively stable across the four sessions, at least for the Continuously-Consistent and Adjusted-Consistent target trials.

Degree of Consistency Phase, RT--Reversal Trials and Supplemental-VM Condition. During Phase 2, a category used

Table 33. Mean Reaction Time and Difference Scores for the Continuously-Consistent and Adjusted-Consistent Categories by Degree-Group, Across the Degree of Consistency Phase

	Session			
	7	8	9	10
Degree-Group 100				
Continuously-Consistent	608	600	592	556
Adjusted-Consistent	592	580	569	549
Difference	16	20	23	7
Degree-Group 67				
Continuously-Consistent	609	581	559	576
Adjusted-Consistent	671	656	637	656
Difference	-62	-75	-78	-80
Degree-Group 50				
Continuously-Consistent	591	587	558	556
Adjusted-Consistent	654	601	594	582
Difference	-63	-14	-36	-26
Degree-Group 33				
Continuously-Consistent	609	579	562	534
Adjusted-Consistent	724	680	653	660
Difference	-115	-101	-91	-126

in the Adjusted-Consistent condition also served as a distractor on some trials for Degree-Group 67, Degree-Group 50, and Degree-Group 33. This manipulation is referred to as a half-reversal condition (Rogers, 1989) because only the role of the previous CM target is reversed. (When both CM targets and CM distractors reverse roles, the condition is referred to as a full reversal.) The half reversal trials and Supplemental-VM trials were examined across the four sessions of the Degree of Consistency Phase. Mean, correct-trial RT and accuracy for each Training Category in each session are presented by Degree-Group in Table 34. A Degree-Group X Training Category (half-reversal and Supplemental-VM) X Practice (3 X 2 X 4) ANOVA was conducted on the correct-trial RTs from the half-reversal and Supplemental-VM trials. (Degree-Group 100, which had no half-reversal trials, was excluded from this analysis.) Only the effect of Practice reached significance ($\underline{F}(3,63) = 19.79$). Although both conditions improved with practice, there was no evidence that the Supplemental-VM and half-reversal trials differed ($\underline{F}(1,21) = 2.73$). The Training Category did not interact with practice ($\underline{F}(3,63) = 1.06$), or Degree-Group ($\underline{F}(2,21) = .77$). Hence, we conclude that the reversal trials resulted in "novice-level" performance which is expected for half-reversal conditions (Rogers, 1989).

The initial reversal effect, $(1 - ((\text{Continuously-Consistent}_{rt} - \text{Reversal}_{rt})/\text{Continuously-Consistent}_{rt}))$, was -29, -23, and -23 percent for Degree-Group 67, Degree-Group 50, and Degree-Group 33 reversal trials, respectively. These reversal effects are consistent with the data reported by Rogers (1989) and Dumais (1979) for half-reversal conditions. After four practice sessions, the reversal effect was still present for Degree-Group 67; however, the effect of reversal had diminished to 12 and 10 percent for Degree-Group 50 and Degree-Group 33, respectively. The reversal effects suggest that an "automatic process,"

Table 34. Mean Reaction Time and Percent Accuracy for Reversal and Supplemental-VM Trials as a Function of Degree-Group and Sessions of Practice (Degree of Consistency Phase)

		Session			
Reversals		7	8	9	10
Degree-Group 100	RT	-	-	-	-
	Accuracy	-	-	-	-
Degree-Group 67	RT	787	735	719	757
	Accuracy	89	91	91	89
Degree-Group 50	RT	727	675	678	664
	Accuracy	91	93	92	93
Degree-Group 33	RT	747	725	687	699
	Accuracy	92	92	92	93

		Session			
Supplemental-VM Trials		7	8	9	10
Degree-Group 100	RT	756	698	691	660
	Accuracy	93	93	94	93
Degree-Group 67	RT	795	737	722	725
	Accuracy	92	93	93	93
Degree-Group 50	RT	723	679	664	660
	Accuracy	93	93	93	93
Degree-Group 33	RT	727	702	675	668
	Accuracy	92	94	93	94

developed during the Training Phase, was present at the beginning of the Degree of Consistency Phase for all groups. Further, the reversal-effect estimates suggest that the automatic process was still present at the end of the Degree of Consistency Phase for Degree-Group 67. However, the reversal effect late in Phase 2 was diminished for Degree-Group 50 and Degree-Group 33, suggesting that the 50 and 33-percent degree-of-consistency manipulation had led to a weakening of the previously developed automatic process.

Degree of Consistency Phase, Accuracy--Reversal Trials and Supplemental-VM Condition. A Degree-Group X Training Category X Practice (3 X 2 X 4) ANOVA was also conducted on the accuracy data for the half-reversal and Supplemental-VM conditions. A significant Degree-Group X Training Category interaction ($\underline{F}(2,21) = 6.01$) as well as significant main effects of Training Category ($\underline{F}(1,21) = 47.92$) and Practice ($\underline{F}(3,63) = 6.01$) were found. Inspection of the means in Table 34 shows that Degree-Group 67 was the least accurate on reversal trials; hence, the reversal effects reported above for Degree-Group 67 are most likely underestimated relative to Degree-Group 50 and Degree-Group 33.

Degree of Consistency Phase, Summary. Manipulating the degree of consistency and changing the distractor categories disrupted performance. RTs increased for all conditions, but that increase was influenced by the degree-of-consistency manipulation. Performance in the Adjusted-Consistent condition was worse at lower levels of consistency. The pattern of results produced by the degree-of-consistency manipulation remained stable across the four sessions of Phase 2. In addition, half-reversal trials showed performance equivalent to Supplemental-VM performance across the four sessions of the Degree of Consistency Phase. The reversal effects were strong and in line with previous studies examining half-reversal effects in the first session

of this phase. By the end of the Degree of Consistency Phase, only Degree-Group 67 demonstrated a strong half-reversal effect. The subjects in Degree-Group 50 and Degree-Group 33 showed an attenuated reversal effect within the last session of the Degree of Consistency Phase.

Retraining Phase

Reaction Time. The mean, correct-trial RT data are presented in Table 35 as a function of Degree-Group, Training Category (Continuously-Consistent, Adjusted-Consistent, and New CM), and sessions of retraining. A Degree-Group X Training Category X Practice (4 X 3 X 4) ANOVA revealed a significant interaction of Training Category and Practice ($F(6,164) = 11.22$). No other comparisons reached significance. The significant Training Category X Practice interaction appears to be due to the fact that the New CM condition was slower at the first session of retraining and improved more during retraining than the other two training categories (New CM improved to a level of performance close to the other two Training Category conditions). This finding indicates that subjects benefitted from their prior consistent training on the Continuously-Consistent and Adjusted-Consistent categories, and that the degree-of-consistency manipulation did not completely disrupt the automatic process developed in the Training Phase. However, a statement that no disruption occurred due to the degree-of-consistency manipulation seems premature (as we explain below).

Retraining Phase - Accuracy. The accuracy data, presented as a function of Training Category by Degree-Group and Practice, are provided in Table 36. As might be expected from examining that table, there were no significant differences in accuracy among the conditions. An ANOVA (Degree-Group X Training Category X Practice) found that none of the main effects were significant: Degree-

Table 35. Mean Reaction Time for Each Retraining Category in Each Session of the Retraining Phase, as a Function of Degree-Group

	Degree-Group	Session 11	12	13	14
Continuously-Consistent	100	553	529	519	509
	67	549	513	514	491
	50	542	524	514	507
	33	549	533	518	517
Adjusted-Consistent	100	529	508	498	482
	67	542	511	508	493
	50	539	517	488	484
	33	567	534	514	520
New CM	100	601	550	524	524
	67	619	573	560	541
	50	614	565	545	536
	33	622	561	547	539

Table 36. Mean Percent Accuracy for Each Retraining Category in Each Session of the Retraining Phase, as a Function of Degree-Group

	Degree-Group	Session 11	12	13	14
Continuously-Consistent	100	93	93	93	92
	67	94	93	93	93
	50	94	94	94	94
	33	93	93	93	94
Adjusted-Consistent	100	93	92	92	92
	67	93	93	93	93
	50	94	94	94	93
	33	93	93	92	93
New CM	100	93	93	94	93
	67	92	94	93	92
	50	93	93	93	93
	33	93	93	93	93

Group, $\underline{F}(3,28) = 2.57$; Training Category, $\underline{F}(2,56) = 1.95$; and Practice, $\underline{F}(3,82) = 1.44$. Neither the Degree-Group X Training Category interaction ($\underline{F}(6,56) = 1.65$) nor the Degree-Group X Practice interaction ($\underline{F}(6,56) = 1.65$) were significant. Finally, the Degree-Group X Training Category X Practice interaction failed to reach significance ($\underline{F} < 1$).

Retraining Phase - Disruption Due to Degree of Consistency. Of particular interest in the Retraining Phase was the level of performance in the Adjusted-Consistent condition when 100-percent consistency was first reinstated. If performance improvement is the result of a target/distractor strengthening differentiation process (Shiffrin and Czerwinski, 1988), then initial retraining performance in the Adjusted-Consistent condition (relative to end-of-training performance) should be disrupted as a function of Degree-Group.

The data for the previous consistent search category and the previous degree of consistency category are presented in Table 37, as well as the mean RTs for each condition at the end of training and in the first session of the Retraining Phase. Also presented are the average difference scores which represent the change in performance between the end of training and beginning of retraining for each condition. Given the fact that at the end of the Training Phase the Adjusted-Consistent condition was faster than the Continuously-Consistent condition, a direct comparison between RTs of the Continuously-Consistent and the Adjusted-Consistent conditions could be misleading (e.g., Cook and Campbell, 1979, Chapter 3).

The difference scores, however, do provide a measure of initial performance change in the Retraining Phase relative to each condition's performance at the end of training. The difference between performance at end of training and beginning of retraining is useful for assessing disruption

Table 37. Mean Reaction Time and Difference Scores for Continuously-Consistent and Adjusted-Consistent Category by Degree-Group, in the Last Session of the Training Phase and the First Session of the Retraining Phase

	Degree-Group	Session 6	Session 11	Difference
Continuously-	100	523	553	-30
Consistent	67	515	549	-34
	50	520	542	-22
	33	519	549	-30

	Degree-Group	Session 6	Session 11	Difference
Adjusted-	100	502	529	-27
Consistent	67	504	542	-38
	50	508	539	-31
	33	500	567	-67

in the Adjusted-Consistent condition, given that a corresponding measure is available for the Continuously-Consistent condition. In combination, these scores provide a measure of the effect of simply changing distractors (difference scores for the Continuously-Consistent condition) and the combined effect of the degree-of-consistency manipulation and the change in the distractors (difference scores for the Adjusted-Consistent condition).

Because of violations of parametric assumptions (e.g., homogeneity of variance and skewness of the distributions), the difference-score data could not be analyzed using an ANOVA. Hence, utilizing the difference scores, we examined the level of disruption for the Continuously-Consistent condition relative to the disruption in the Adjusted-Consistent condition for each subject within each Degree-Group. That is, we determined, for each subject in a given Degree-Group, whether the Adjusted-Consistent condition was more or less disrupted than the Continuously-Consistent condition. Seven of the eight subjects in Degree-Group 33 showed more disruption in the Adjusted-Consistent condition than the Continuously-Consistent condition. Four of the eight subjects participating in Degree-Group 50 and four of the eight subjects in Degree-Group 67 were more disrupted in the Adjusted-Consistent condition than the Continuously-Consistent condition. Only two of the eight subjects participating in Degree-Group 100 were more disrupted in the Adjusted-Consistent condition compared with the Continuously-Consistent condition. Fisher's exact probability test (Siegel, 1956, p. 96-104) showed that the probability of Degree-Group 100 and Degree-Group 33 being classified as having the same level of disruption is less than 0.02. The null hypothesis under this test could not be rejected for comparisons between Degree-Group 100 and Degree-Group 67 as well as between Degree-Group 100 and Degree-Group 50 ($p > .24$ in both cases). For Degree-Group

33, even the less powerful sign test demonstrated that greater disruption (relative to performance at the end of the Training Phase) occurred for the Adjusted-Consistent condition compared to the corresponding Continuously-Consistent condition ($p < .035$). Consistent with the Fisher exact probability test, the sign test failed to reveal differences in disruption between the Adjusted-Consistent and Continuously-Consistent conditions for Degree-Group 100, Degree-Group 67, and Degree-Group 50.

For each Degree-Group, we also examined the percentage of disruption for each Adjusted-Consistent condition <u>relative</u> to each group's corresponding performance in the Continuously-Consistent condition using the formula presented above in the report of the Degree of Consistency Phase. The relative disruption in performance for the Adjusted-Consistent conditions was -11 percent (negative implies improvement), 10.5, 29, and 55 percent for Degree-Group 100, Degree-Group 67, Degree-Group 50, and Degree-Group 33, respectively. These data are in line with the difference scores.

<u>Summary</u>. In the Retraining Phase, both the Continuously-Consistent and the Adjusted-Consistent conditions were superior to the New CM condition, regardless of Degree-Group participation. Hence, we must conclude that the benefit of the consistent training received during the Training Phase was not eliminated during the Degree of Consistency Phase even for Degree-Group 33. However, more detailed analyses showed that the Adjusted-Consistent category of Degree-Group 33 was more disrupted during the Degree of Consistency Phase than the Adjusted-Consistent categories of the other Degree-Groups.

Discussion

The subjects in all Degree-Groups seemed to develop skilled visual search during the Training Phase. Their performance improved during training with the performance-practice function fitting a general power function. Performance improvement, well-described by the "ubiquitous law" of practice (Newel and Rosenbloom, 1981), is one indication that automatic processing had developed by the end of 6000 practice trials. In the Degree of Consistency Phase, performance on reversal trials indicated substantial disruption--another indication of automatic process development.

The pattern of performance, between and within Degree-Groups during the Degree of Consistency Phase, allows us to answer, to some level, the questions posed in the introduction of this section. The disruption on Adjusted-Consistent condition target trials was a function of Degree-Group, i.e., more disruption occurred as consistency decreased. Detection of targets from the Continuously-Consistent condition was slower in Phase 2, but the slowing was unrelated to Degree-Group (no statistically significant differences among Degree-Groups in the Continuously-Consistent condition). The disruption of performance in the Continuously-Consistent condition was most likely due to the change from consistent distractors to new, VM distractors.[15] Hence, because the Adjusted-Consistent conditions were disrupted differentially as a function of degree of consistency and the Continuously-Consistent condition showed minimal, uniform disruption, we can conclude that the

[15] With practice, CM distractors are weakened and CM target stimuli are strengthened (Dumais, 1979; Rogers, 1989; Shiffrin and Czerwinski, 1988). Novel distractors have a higher strength level than well-trained CM distractors (Dumais, 1979; Rogers, 1989). Changing the target-distractor strength ratio can disrupt performance (Dumais, 1979; Rogers, 1989; Shiffrin and Czerwinski, 1988).

differential disruption of (or the differential need to inhibit) one automatic process does not necessarily differentially affect automatic processing on other tasks. This conclusion may not apply beyond situations such as those used in the present experiment where the stimuli triggering automatic processes are segregated. However, the present finding is still important because it suggests that (1) changes in task context will not necessarily disrupt automatic processes and (2) inhibiting one automatic process does not necessarily lead to an inability to "let go" (Schneider and Fisk, 1983) of other automatic processes.

What is the locus of the differential disruption effects for the target trials in the Adjusted-Consistent condition found across the Degree-Groups? Recall that the differential disruption for those trials occurred immediately in the first session of the Degree of Consistency Phase. If weakening of the automatic process were the sole source of that disruption, we would expect neither such substantial disruption nor a differential disruption among the Degree-Groups to occur so quickly. Subjects received 3000 target-strengthening trials before transferring to the Degree of Consistency Phase. The first session of the Degree of Consistency Phase included 200 target-present trials (strengthening trials) and 100, 200, and 400 reversal trials (target-weakening trials) for Degree-Group 67, Degree-Group 50, and Degree-Group 33, respectively. There were a minimal number of "weakening" trials to produce such a substantial differentiation among the Degree-Groups in the target trials of the Adjusted-Consistent condition.

Although there is some evidence that differential weakening was occurring, the major source of the differential disruption of the target trials in the Adjusted-Consistent condition most likely comes from the

need to inhibit automatic processing more often as degree of consistency decreased. Recall that the Adjusted-Consistent condition contained the same number of total trials and the same number of target trials within a block. However, by design there were more reversal trials as degree of consistency decreased. Hence, the probability of having a target trial (where an automatic process could be used) follow a string of reversal trials (where an automatic process must be inhibited) increased as degree of consistency decreased. (For example, a subject in Degree-Group 33 might encounter an Adjusted-Consistent target trial preceded by three reversal trials and followed by three reversal trials. In Degree-Group 67, the likelihood of such strings of trials is decreased.) As the need to inhibit the automatic process increases (and therefore the overall reliability of the output of the automatic process decreases), subjects are more likely to rely on the slower but more reliable, controlled process (Schneider and Fisk, 1983). As the ratio of target to reversal trials increases, subjects will rely less on the output of an automatic process by "rechecking" its output with the output of the control process (see Schneider and Fisk, 1982b, 1983 for related discussions). Such a change in strategy would account for the almost immediate degree-of-consistency effect.

The potential change in strategy notwithstanding, is there evidence that weakening is occurring? Both the reversal effects within the Degree of Consistency Phase and the disruption when 100-percent consistency is reinstated (Retraining Phase) suggest that some weakening is occurring; however, we are quick to add that weakening is minimal even for Degree-Group 33. Subjects in Degree-Group 67, Degree-Group 50, and Degree-Group 33 showed a strong reversal effect in the first session of the Degree of Consistency phase. This is expected if an automatic process was

developed during the Training Phase. After four sessions in the Degree of Consistency Phase, the reversal effect was still strong for Degree-Group 67, less strong for Degree-Group 50, and weakest for Degree-Group 33.

When transferring to the Retraining Phase, in which 100-percent consistency was restored, the performance of all Degree-Groups in the Adjusted-Consistent condition was better than their performance in the New CM condition. Hence, even Degree-Group 33 retained some level of the automatic process developed in the Training phase. Yet, more detailed analysis of the data indicated that Degree-Group 33 was more disrupted in the Adjusted-Consistent condition, relative to the Continuously-Consistent condition. This is further evidence that some, although minimal, disruption of the automatic process had occurred during the Degree of Consistency Phase.

It is perhaps not surprising that the disruption to the automatic process, even for Degree-Group 33, would be minimal. The number of target-weakening trials was relatively small compared to the number of target strengthening trials for the Adjusted-Consistent condition. Recall that all subjects received 3000 target-strengthening trials in the Adjusted-Consistent condition during the Training Phase. During the Degree of Consistency Phase, all subjects received an additional 800 target-strengthening trials for stimuli in that condition. In the first session of the Retraining Phase subjects received an additional 350 Adjusted-Consistent condition target strengthening trials for a total of 4150 target-strengthening trials across the phases of the experiment. The total number of weakening (reversal) trials was 400, 800, and 1600, respectively, for Degree-Group 67, Degree-Group 50, and Degree-Group 33.

Some data exist that suggest several thousand reversal trials may be required before a well-learned automatic

process is "unlearned" (Fisk, Lee, and Rogers, in press; Shiffrin and Schneider, 1977). Therefore, it is not surprising that performance in the first session of the Retraining Phase in the Adjusted-Consistent condition was superior to the New CM condition--even for Degree-Group 33. The automatic process was still sufficiently strong to benefit performance, relative to processing new CM stimuli, even in Degree-Group 33 given the change in context such that the need to inhibit action based on the output of the automatic process was removed. The Retraining Phase data suggest that disruption of the automatic process will be minimal, even when 67 percent of the occurrence of the well-learned stimuli require inhibition of the automatic process and the encounter with such inconsistency is not prolonged. Given a more sensitive procedure, we may have been able to demonstrate a more graded effect of degree of inconsistency on the disruption of the automatic process; however, the disruption would be minimal nonetheless.

In addition to value from a theoretical perspective, the present findings have practical implications. First, it appears that inhibiting one automatic process will not dramatically effect a different automatic process as long as both automatic processes are independent. This finding is important because it suggests that part-task training can be developed to retrain one automatic process without interfering with other related but independent automatic processes.

The present data also suggest that, assuming personnel are well-trained, if those individuals encounter inconsistency in the operational environment the well-learned automatic process will survive--at least within the limits presently tested. Such a finding has both positive and negative aspects from an operational perspective.

The positive aspect is that skill will survive if one task must be performed that requires a process incompatible with a well-learned task component, at least if performance of the incompatible task is not prolonged. However, the negative aspect is that if a new, incompatible task component is to replace the old automatic component, training time for the new task will be long. In addition, the present data suggest that if an incompatible task must be performed along with the well-learned automatic process (i.e., both tasks must be performed together), performance will be disrupted on both tasks for quite some time. The well-learned automatic component will be better than novice-level performance, but it will be worse than highly skilled performance. Performance of the task incorporating the component that is incompatible with the well-learned automatic process will remain at or be worse than novice-level performance (depending on how incompatibility is introduced--full vs. half reversals) for quite some time.

V. EXPERIMENTAL SERIES 4: IMPROVEMENT IN VISUAL SEARCH IN TRAINING ENVIRONMENTS OF VARYING DEGREES OF INCONSISTENCY

Introduction

Performance improvement in visual search seems to be the result of various learning mechanisms (e.g., Czerwinski, 1988; Rogers, 1991). Such improvement seems to be the result of learning general search strategies; devising optimal, stimulus-specific search strategies; and the developing of automatic processing (automatic attention attraction). We have suggested, in other sections of this report (e.g., Section II) that an individual's performance improvement is guided by the same factors (learning general search strategies; then optimal, stimulus-specific search strategies) early in practice for both CM and VM training conditions. Qualitative differences between CM and VM performance are seen late in practice and generally when the implementation of the attentive optimal search is difficult (see Czerwinski, 1988 for related views).

Fisher (1986; Fisher and Tanner, in press) provides a formal model that describes the development of optimal feature search. This model assumes that individuals learn an optimal set of feature comparisons for particular target-distractor combinations. Selection of the critical features and the optimal search sequence requires experience with a given target-distractor set. Hence, as long as critical features can be identified, learning is predicted to occur in both traditional CM and VM training procedures. According to Fisher's theory, the difference between CM and VM performance occurs because typically, there are fewer combinations of target-distractor features to learn in CM than VM. If the number of target-distractor combinations are equated (or more VM training is provided), some evidence exists that performance can be equated for CM and VM

conditions (Fisher, 1986); however, the underlying learning seems to be different (Czerwinski, 1988).

It seems clear that development of an optimal search strategy is critical for skilled visual search. However, there is some indication that the training environment interacts with amount of practice in regards to development of optimal search strategies. We do know that the distribution of practice on particular target-distractor pairings affects the development of optimal search. Lee, Rogers, and Fisk (1991) presented results of a study that examined the distribution of practice across target-distractor pairings for pure CM, pure VM, and "cycle" conditions. The cycle conditions maintained a constant target-distractor pairing (e.g., stimulus set A was always a target set when stimulus set B was a distractor set) but the stimuli served as both targets and distractors (e.g., set A was a target set when set B was a distractor set, set B was a target set when set C was a distractor set, set C was a target set when set A was a distractor set). Hence, optimal search could more efficiently develop in the "cycle" conditions compared with pure VM; yet, because the stimuli were not consistently mapped as targets or distractors, an automatic process could not develop. The study showed that if subjects received at least ten repetitions of a given condition, optimal search strategies could develop. When repetitions were only five or fewer, development of optimal search strategies seemed to be disrupted.

The Lee et al. (1991) study notwithstanding, at present we know little about how the development of optimal search strategies are affected by training environments. Hence, the present study examines another training situation to learn more about factors affecting the development of optimal search.

Method

<u>Subjects</u>. Seventeen students from a southeastern university (ten males and seven females) participated in this experiment. Participants ranged from 17- to 26-years-old with an average age of 19.35. Students received course credit or $4.00 per session for their time. All participants had visual acuity of at least 20/30 (far vision) and 20/40 (near vision).

<u>Stimuli</u>. The stimuli for the present experiment were 11 semantically unrelated categories as determined by the Collen, Wickens, and Daniele (1975) norms. The categories were: Alcohol, Clothing, Earth Formations, Furniture, Occupations, Relatives, Trees, Units of Time, Vegetables, Vehicles, and Weapons. Six high associates from each category (Battig and Montague, 1969), four to seven letters long, were chosen as exemplars.

A calibration study was conducted to find a set of relatively equally confusable stimuli. Based on that study, the categories Furniture, Vehicles, and Trees were chosen as target categories. The assignment of categories to search conditions was counterbalanced using a partial Latin Square.

<u>Apparatus</u>. Microcomputers were programmed to control the timing of the displays, stimuli presentation, and response collection using Psychological Software Tools' Micro Experimental Laboratory software (Schneider, 1988). The data were collected using three EPSON Equity I+ microcomputers with EPSON MBM 2095-5 green monochrome monitors. The EPSON Q-203A keyboard was altered by exchanging the "7," "4," and "1" numeric keypad keys with the "T," "M," and "B" keys, respectively. For all experimental sessions, pink noise played at approximately 57db(A) to attenuate background noise. All subjects worked

in the same room at individual, partitioned workstations and were monitored by a laboratory assistant.

Procedure. During the first session, subjects were given instructions and received 30 CM trials as orientation to the task. These orientation trials allowed the subjects to become familiar with the experimental protocol. The categories used for the orientation trials (Colors and Birds) were not used in the remainder of the experiment. Subjects began the first experimental session immediately following orientation.

For each trial, the memory set contained one category label (memory-set size of one) and the display contained three words (display-set size of three). An individual trial consisted of the following sequence of events. The subject was presented with one category label as the memory-set item which he/she was allowed to study for a maximum of 20 seconds. Subjects were instructed to press the space bar to initiate the trial. Three plus signs were then presented for 0.5 seconds in the center of the screen to allow the subject to localize his/her gaze. A target (i.e., an exemplar from the memory-set category) was present for every trial. The display set consisted of three words (the target word and two distractor words) presented in a column. The subject's task was to indicate the location of the target (i.e., top, middle, or bottom) by pressing the corresponding key (labeled "T," "M," or "B").

Feedback. Participants received performance feedback at three distinct times: (1) at the end of each trial, (2) the end of each block of 50 trials, and (3) the start of each session. After correct trials, RT was displayed in milliseconds. After incorrect trials, an error tone sounded, followed by a display of the correct response. At the end of each block (50 trials), the subject's average RT and accuracy for that block were presented.

At the end of every block of trials, subjects received feedback concerning their accuracy. If a subject's mean accuracy (for that block) fell below 92 percent, a message encouraging him/her to respond more carefully on the next occurrence of that same type of block was displayed. If the subject's mean accuracy exceeded 96 percent, a message was displayed encouraging him/her to respond faster on the next occurrence of that same type of block. If the subject received either message, he or she was also reminded at the start of the next block of that specific type of trials with the message, "Remember to respond [carefully or faster] on this group of trials."

Subjects were encouraged to respond as fast as possible, while keeping their accuracy within the range described above. Before each session, participants privately received feedback on their previous day's performance. Each subject was shown his/her RT, and encouraged to improve his/her performance from session to session.

Experimental Sequence

The experiment was run in two phases. The first was a Degree of Consistency Training Phase, designed to examine the effects of training in visual search at various levels of consistency. The second phase, pure CM training, is referred to as the Consistent Training Phase. This phase examined the effect of the previous degree of consistency manipulation on subsequent development of skilled performance. The dependent measures in both phases were RT and accuracy. A summary of the experimental sequence and the conditions within each phase is presented in Table 38.

Table 38. Training and Transfer Condition Summary: Section IV

Degree of Consistency Training Phase		Consistent Training Phase	
Training Category	# of Trials[a]	Training Category	# of Trials
<u>Degree-Group 100</u>			
Continuously-Consistent	200	Continuously-Consistent	350
Adjusted-Consistent	200	Adjusted-Consistent	350
VM	600	New CM	350
<u>Degree-Group 67</u>			
Continuously-Consistent	200	Continuously-Consistent	350
Adjusted-Consistent	200	Adjusted-Consistent	350
Reversal[b]	100	New CM	350
VM	500		
<u>Degree-Group 33</u>			
Continuously-Consistent	200	Continuously-Consistent	350
Adjusted-Consistent	200	Adjusted-Consistent	350
Reversal	400	New CM	350
VM	200		

[a]The number of trials column indicates the number of trials within one session.
[b]Reversal refers to those trials for which a word from the Adjusted-Consistent category serves as a distractor item.

Degree of Consistency Training Phase, Experimental Design

In the Degree of Consistency Training Phase, all 17 subjects received training on two of the three possible target categories (Furniture, Vehicles, and Trees). (The remaining target category was used as a new CM target category in the Consistent Training Phase, described below.) One of the two target categories is referred to as the Continuously-Consistent category because this category was consistently mapped as a target throughout the entire experiment (both phases). This category was included as a within-subject control category. The other target category is referred to as the Adjusted-Consistent category because its consistency was adjusted, or changed, as a function of the phase of the experiment and the Degree-Group assignment (see below). The consistency of the Adjusted-Consistent category was either 100, 67, or 33 percent during the Degree of Consistency Training Phase (adjusted as a function of Degree-Group), and 100 percent during the final phase. (See Table 39 for a summary of the Training Category consistency as a function of experiment phase and Degree-Group.)

Each Degree-Group also received VM training in the Degree of Consistency Training Phase. The four categories that made up the VM condition served as both targets and distractors. The four VM categories served as distractors for the Continuous-Consistent condition and for the Adjusted-Consistent condition. The VM categories served as targets for some trials in the Adjusted-Consistent condition blocks (except for Degree-Group 100) and during the VM trial blocks.

The Degree of Consistency Training Phase included six 1000-trial sessions (a total of 6000 training trials). The 17 subjects were randomly assigned to one of the three between-subjects groups (Degree-Group). The Degree-Groups

Table 39. Percent Consistency for Each Training Category as a Function of Degree-Group for Each Phase of the Experiment

	Phase of the Experiment	
Degree-Group	Degree of Consistency Training	Consistent Training
Degree-Group 100		
Continuously-Consistent	100	100
Adjusted-Consistent	100	100
VM	23	--
New CM	--	100
Degree-Group 67		
Continuously-Consistent	100	100
Adjusted-Consistent	67	100
VM	25	--
New CM	--	100
Degree-Group 33		
Continuously-Consistent	100	100
Adjusted-Consistent	33	100
VM	33	--
New CM	--	100

were: Degree-Group 100 (six subjects), Degree-Group 67 (five subjects), and Degree-Group 33 (six subjects).

In the Degree of Consistency Training Phase, the between-subject independent variable was Degree-Group (Degree-Group 100, Degree-Group 67, and Degree-Group 33);this refers to the consistency of the Adjusted-Consistent category. The within-subject independent variables were Training Category and Practice (six sessions of practice). The Training Categories were presented in randomly ordered blocks (50 trials per block). There were three Training Category levels: (1) Continuously-Consistent, (2) Adjusted-Consistent, and (3) VM. (Degree-Group 67 and Degree-Group 33 also had "reversal" trials, which are explained below.)

<u>Continuously-Consistent Condition (Within-Subject Control Condition)</u>. The Continuously-Consistent category served as a within-subject control. There were four blocks of the Continuously-Consistent condition (a total of 200 trials) in each session. On a given Continuously-Consistent condition trial, distractors were chosen at random from two of the four VM categories.

<u>Adjusted-Consistent Condition (Degree of Consistency Manipulation)</u>. In Phase 2 of the experiment, the Adjusted-Consistent condition underwent a degree-of-consistency manipulation. The degree of consistency was manipulated between subjects and was either 100, 67, or 33-percent consistent. The degree of consistency of the Adjusted-Consistent category was determined by the Degree-Group to which a subject was assigned.

Degree of consistency of the Adjusted-Consistent category was determined by varying the ratio of the number of times category words appeared as targets versus distractors. The frequency with which the category appeared

as a target was equal, whereas the frequency of its appearance as a distractor was varied. For each Degree-Group, there were 200 target trials of the Adjusted-Consistent condition per session. The number of times the Adjusted-Consistent category served as a distractor was varied across Degree-Groups. Let (t:d) represent the number of times the Adjusted-Consistent category appeared as a target (t) versus a distractor (d). For Degree-Group 100, the target/distractor ratio was 200:0 per session. For Degree-Group 67 and Degree-Group 33, the target/distractor ratios were 200:100 and 200:400 respectively. The trials for which a word from the Adjusted-Consistent category occurred as a distractor are referred to as "Reversal Trials" (these trials are actually "half-reversal" trials, see Dumais, 1979; Rogers, 1989).

Within a block of Adjusted-Consistent trials, the Adjusted-Consistent category served either as a target, or a distractor on a particular trial (as defined above). When the Adjusted-Consistent category served as a target, two distractor words were chosen at random--one each from two of the VM categories. When a word from the Adjusted-Consistent category appeared as a distractor, the memory set contained a VM category. The other distractor for that trial was randomly chosen from one of the three remaining VM categories.

Degree of consistency was defined across a session; however, to create the appropriate degree of consistency in each block, it was necessary to either match or closely approximate the overall degree of consistency. For each block of trials for Degree-Group 100, the Adjusted-Consistent category served as a target in 50 trials (100 percent). For a given block of trials, the overall degree of consistency could only be approximated for Degree-Group 67 and Degree-Group 33. Therefore, in each block of

Adjusted-Consistent trials, the degree-of-consistency category served as a target in at least 30 trials and as a distractor in at least 15 trials for Degree-Group 67. For Degree-Group 33, the Adjusted-Consistent category served as a target in at least 15 trials and as a distractor in at least 30. For the remaining five trials in each block, whether the Adjusted-Consistent category served as target or as a distractor was determined randomly. The only constraint was that the Adjusted-Consistent category serve as a target 200 total times per session and as a distractor either 100 or 400 times per session, for Degree-Group 67 or Degree-Group 33, respectively.

To create the appropriate degree of consistency, it was also necessary to vary the number of Adjusted-Consistent trial blocks across Degree-Group. Subjects participating in Degree-Group 100, Degree-Group 67, and Degree-Group 33 received four, six, and 12 blocks of Adjusted-Consistent trials, respectively. Because all Degree-Groups received an equal number of total trials per session, each Degree-Group also received a different number of VM trial blocks (see below).

Varied Mapping Search Condition

In each session of the Degree of Consistency Phase, subjects performed either 12, ten, or four blocks of VM search (600, 500, or 200 total VM trials per session, for Degree-Group 100, Degree-Group 67, or Degree-Group 33, respectively).

In the blocks of VM trials, the four VM categories served as both targets and distractors. When a word from any one VM category was the target word, two distractor words were chosen at random--one each from two of the remaining VM categories. Within the VM condition, each VM category served as a target category 150, 125, or 50 times

and as a distractor category 300, 250, or 100 times for Degree-Group 100, Degree-Group 67, or Degree-Group 33, respectively. However, because the VM items were used as distractors for the Continuously-Consistent blocks and the Adjusted-Consistent blocks (as well as targets for some trials within the Adjusted-Consistent blocks), the actual degree of consistency of the VM items was 0.23, 0.25, and 0.33 for Degree-Group 100, Degree-Group 67, and Degree-Group 33, respectively.

Consistent Training Phase, Experimental Design

In Phase 2, the Adjusted-Consistent category acted only as a target category. Degree of consistency (the between-subject variable from the Degree of Consistency Training Phase) was not manipulated in the Consistent Training Phase; hence, each Degree-Group received CM training on all target categories.

Subjects completed four sessions in the Consistent Training Phase (1050 trials per session). Each session consisted of 21 blocks of 50 trials. The Training Category variable now consisted of: (1) the Continuously-Consistent condition, (2) the Adjusted-Consistent condition, and (3) a New CM condition (the category used in the New CM condition was not used in the first phase of the experiment). There were seven blocks of each Training Category per session with presentation order randomly determined. Two new categories made up the distractor set. On each trial, the display set consisted of a target word and two distractor words (one distractor chosen at random from each distractor category).

To summarize the design for the Consistent Training Phase, the within-subject independent variables were Training Category (Continuously-Consistent, Adjusted-Consistent, and New CM) and Practice (four CM practice sessions). The between-subject independent variable was

Degree-Group (Degree-Group 100, Degree-Group 67, and Degree-Group 33). However, degree of consistency was not manipulated in this phase.

Results

Degree of Consistency Training Phase. Correct trial RTs for each condition across all sessions of practice are presented in Table 40. The overall Degree-Group X Training Category (Continuously-Consistent, Adjusted-Consistent, and VM) X Practice ANOVA (corrected for unequal subjects) revealed significant main effects of Training Category ($F(2, 28) = 16.90$) and Practice ($F(5, 70) = 38.05$). The interaction between Training Category and Degree-Group was marginally significant ($F(4, 28) = 2.64$, $p < .0551$), and the interaction between Training Category and Practice was significant ($F(10,140) = 2.44$). As suggested by the ANOVA, VM performance improved less than the other two Training Categories and VM performance was not different among the Degree-Groups.

The source of the Training Category X Degree-Group interaction was due to differences among the Adjusted-Consistent condition across Degree-Groups. This is supported by a Degree-Group X Training Category X Practice ANOVA (examining only the Continuously-Consistent and Adjusted-Consistent conditions). For this ANOVA there was a significant interaction between Training Category and Degree-Group ($F(2,14) = 7.60$); however, none of the interactions involving practice reached significance (Fs < 1).

Averaged accuracy data are presented in Table 41 as a function of Degree-Group, Training Category, and Practice. As can be noted from this table, accuracy fluctuated throughout practice; yet, except for the reversal trials, the differences among conditions were minimal. Hence, we

Table 40. Mean Reaction Time (in milliseconds) for Each Training Category by Degree-Group in Sessions 1 Through 6 (Degree Training Phase)

Training Category	Session					
Continuously-Consistent	1	2	3	4	5	6
Degree-Group 100	790	704	644	641	630	604
Degree-Group 67	760	642	629	583	575	586
Degree-Group 33	795	712	685	663	644	643
Adjusted-Consistent	1	2	3	4	5	6
Degree-Group 100	753	692	638	625	616	602
Degree-Group 67	825	681	640	605	605	635
Degree-Group 33	835	720	720	696	668	698
Half-Reversal	1	2	3	4	5	6
Degree-Group 100	--	--	--	--	--	--
Degree-Group 67	833	737	676	676	646	678
Degree-Group 33	828	735	734	723	710	741
VM	1	2	3	4	5	6
Degree-Group 100	819	789	737	739	744	692
Degree-Group 67	822	728	682	644	634	652
Degree-Group 33	822	735	728	734	698	715

Table 41. Mean Percent Accuracy for Each Training Category by Degree-Group in Sessions 1 Through 6 (Degree Training Phase)

Training Category	Session					
Continuously-Consistent	1	2	3	4	5	6
Degree-Group 100	94	94	94	95	95	95
Degree-Group 67	95	95	95	94	94	94
Degree-Group 33	95	94	95	94	94	95
Adjusted-Consistent	1	2	3	4	5	6
Degree-Group 100	94	95	94	95	94	95
Degree-Group 67	95	95	93	95	95	94
Degree-Group 33	96	94	93	93	95	94
Half-Reversal	1	2	3	4	5	6
Degree-Group 100	--	--	--	--	--	--
Degree-Group 67	93	95	94	92	92	91
Degree-Group 33	94	94	93	93	92	93
VM	1	2	3	4	5	6
Degree-Group 100	93	94	93	93	92	93
Degree-Group 67	95	94	94	94	94	93
Degree-Group 33	95	94	94	93	93	94

conclude that accuracy effects do not preclude inferences made about the RT data.

The data presented in Table 40, show that improvement for the Continuously-Consistent condition followed a pattern expected for CM tasks. Although only 1200 trials of practice were provided, practice in the Continuously-Consistent condition resulted in final-level performance which was better than the other conditions. Interestingly, the Continuously-Consistent condition for Degree-Group 33 was slower after training (but not in the first session) than the other Degree-Groups. This suggests that the Degree-Group 33 training environment slightly slowed RT, even for the Continuously-Consistent condition.

Final-level performance on the Adjusted-Consistent condition was as expected, given the degree-of-consistency manipulation. Performance was fastest for Degree-Group 100, somewhat slower for Degree-Group 67, and slowest for Degree-Group 33. In fact, RT on the Adjusted-Consistent condition was about 100 ms slower for Degree-Group 33 than Degree-Group 100. Reversal performance was slower (nonsignificantly) than performance in the VM condition.

VM performance improved for all Degree-Groups; but, the improvement did not reach the level of performance on the Continuously-Consistent condition for any Degree-Group. Improvement in VM did not follow the pattern expected based on Fisher's (1986) feature overlap model. Search strategies appear to be developing for all groups; however, although Degree-Group 33 was the slowest (as predicted), Degree-Group 100 showed only a 23-ms faster RT than Degree-Group 33 in the VM condition. This is surprising because Degree-Group 100 received 3600 trials of VM (828 target trials, 207 target trials per VM category) while Degree-Group 33 received only 1200 VM trials (396 target trials, 99 target trials per VM category). Further problems for Fisher's

theory arise because Degree-Group 67 had the fastest VM performance and that group received 3000 VM trials (750 target trials, approximately 187 target trials per VM category). The ratio of CM to VM trials does not appear to be all the information needed to predict final-level performance in those training conditions. If such information were sufficient to predict performance, we would expect that comparisons between the Continuously-Consistent condition and the VM condition would show the least difference between these conditions for Degree-Group 100, intermediate difference scores for Degree-Group 67, and the largest difference between the Continuously-Consistent condition and VM condition for Degree-Group 33. The difference between the Continuously-Consistent condition and the VM condition was -88 ms, -66 ms, and -72 ms for Degree-Group 100, Degree-Group 67, and Degree-Group 33, respectively.

Consistent Training Phase. Mean, correct-trial RTs for Phase 2 are presented in Table 42; and the corresponding accuracy data are presented in Table 43. A Degree-Group X Training Category (Continuously-Consistent condition, Adjusted-Consistent condition, and New CM) X Practice (3 X 3 X 4) ANOVA was conducted on the RT data from the Consistent Training Phase. There were significant main effects of Training Category ($F(2, 28) = 7.80$) and Practice ($F(3,41) = 18.3$). The interaction between Training Category and Practice also reached significance ($F(6,82) = 6.32$).

The data in Table 42 show that the New CM condition differed from the Continuously-Consistent condition and the Adjusted-Consistent condition when the training environment became totally consistent. In addition, the Adjusted-Consistent condition was slower than the Continuously-Consistent condition, regardless of Degree-Group. By the

Table 42. Mean Reaction Time for Each Training Category by Degree-Group, as a Function of Sessions of Consistent Phase Practice

		Session			
Category	Degree-Group	7	8	9	10
Continuously-Consistent	100	570	550	543	533
	67	560	560	516	513
	33	569	531	550	533
Adjusted-Consistent	100	583	555	539	540
	67	588	566	520	511
	33	586	533	554	536
New CM	100	633	598	566	557
	67	607	602	526	504
	33	648	561	578	562

Table 43. Mean Percent Accuracy for Each Training Category by Degree-Group, as a Function of Sessions of Consistent Phase Practice

Category	Degree-Group	Session 7	8	9	10
Continuously-Consistent	100	94	94	93	93
	67	95	93	93	93
	33	95	95	94	94
Adjusted Consistent	100	94	93	93	93
	67	94	94	94	94
	33	94	94	94	93
New CM	100	94	94	94	92
	67	94	94	95	95
	33	93	94	93	93

end of the Consistent Training Phase, none of the Training Category conditions differed.

An analysis examining performance in the Continuously-Consistent condition and the Adjusted-Consistent condition across Degree-Groups in the last session of the Degree of Consistency Training Phase and the first session of the Consistent Training Phase revealed that the Continuously-Consistent condition was faster than the Adjusted-Consistent condition ($F(2, 14) = 5.09$) However, Degree-Group did not interact with Training Category ($F < 1$). The interaction between Degree-Group and Session (last Degree of Consistency session versus first session of the Consistent Training Phase) was marginally significant ($F(2,14) = 2.77$).

Discussion

The present data replicate the degree-of-consistency effects found in the experiment described in Section IV. RT in the Adjusted-Consistent condition was a direct function of degree of consistency. All subjects exhibited good transfer from the Degree of Consistency Phase to the Consistent Training Phase for the Adjusted-Consistent condition. This finding is somewhat surprising given that the consistency of one Adjusted-Consistent condition was only 33 percent. However, this finding, coupled with the fact that new distractors were used when subjects transferred to the consistent training phase, suggests either: (1) the 1200 target-present trials strengthened the targets in the Adjusted-Consistent condition more than the 2400 reversal trials weakened their attention-calling strength (for Degree-Group 33); or (2) subjects learned optimal search strategies that were target-specific but not distractor-specific.

The latter conclusion seems somewhat unwarranted given the lack of differences among the VM conditions even though

different amounts of VM practice were received across Degree-Groups. If optimal feature search were the primary determinant of performance improvement observed in the Adjusted-Consistent conditions then the expectation would be that the VM conditions would differ and the difference between a given Continuously-Consistent condition and the corresponding VM condition would be a function of the ratio of CM to VM trials.

Schneider and Detweiler (1987) have argued that target strengthening is much faster than distractor weakening. In fact, they have proposed that the difference is up to fourfold. Hence, even with consistency at 33 percent and given the number of target detections relative to reversal trials, one could expect some (but not maximal) target strengthening beyond a "neutral" level of new distractors. Hence, the good transfer to the Consistent Training Phase for the Adjusted-Consistent condition and the small differences between the Adjusted-Consistent conditions and the Continuously-Consistent conditions in the Consistent Training Phase could be attributed to the target:distractor strength ratio reaching some threshold for target stimuli in both the Adjusted-Consistent condition and the Continuously-Consistent condition.

The present data do not argue against the development of optimal search strategies. There is substantial evidence that development of search strategies can be crucial for performance improvement in visual search. However, the present data argue that the training environment may interact with other factors to reduce the development of optimal search strategies. However, the present data add to the list of situations not supportive of an instance-based view of automaticity (Logan, 1988). Such a theoretical perspective would predict poor transfer when switching between phases of the experiment. In fact, both the

Continuously-Consistent condition and the Adjusted-Consistent condition demonstrated good transfer.

VI. EXPERIMENTAL SERIES 5: A COMPARISON OF PART- AND WHOLE-TASK TRAINING PROCEDURES IN THE DEVELOPMENT AND RETENTION OF A HIGH-PERFORMANCE, SKILL-BASED, DECISION-MAKING TASK

Introduction

An important consideration for the assessment of training techniques involves comparing the benefits of part-task and whole-task training in complex skill development. Part-task training refers to a class of procedures by which specific task components are trained, prior to practice on the whole task. A key assumption underlying part-task training is that identification of, and training with, individual task components may improve whole-task performance more effectively than training on the entire task.

To this end, an experiment using a highly complex, dispatching task was conducted. This task is a conceptual analog of the tactical resource allocation required in real-world, battle-management tasks. The experiment was performed to investigate the potential benefits of part-task training the memory components of the dispatching task and to determine the effect of prior knowledge of the whole task on part-task training effects. The dispatching task has several procedural components and requires a substantial amount of declarative knowledge; it is also heavily rule-based.

In Phase 1 of the experiment, high-performance skill development as a function of whole-task versus part-task training (and type of part-task training) was examined during training and transfer. In the second phase, skilled performance on the dispatching task was measured 60 days following the final (transfer) session of practice for all conditions. A brief overview of part-task training procedures and a description of their advantages and

disadvantages are presented in the following sections. An overview of the experiment follows this summary. Finally, a description of the dispatching task employed in this investigation is provided.

Wightman and Lintern (1985) proposed that differential transfer provides an optimal measure of the effectiveness of part-task training. According to Wightman and Lintern, differential transfer refers to the relative effects of equivalent experience with experimental (part-task training) and control (whole-task training) conditions on task performance. If differential transfer is less than 100 percent, it may be concluded that part-task training is less efficient than whole-task training. However, part-task training is successful in developing skills useful for performance of a criterion task. If differential transfer is greater than 100 percent, it may be concluded that part-task training is superior to whole-task training.

Wightman and Lintern (1985) describe three main methods of part-task training: segmentation, fractionation, and simplification. Segmentation involves partitioning the task so that subtasks are practiced separately, then recombined into the whole task. Fractionation is used for tasks in which two or more skill components must be performed simultaneously. Simplification makes a difficult task easier by modifying the characteristics of the task. This training method is related to adaptive training; both procedures involve simplification of the whole task rather than decomposition and separate training of subcomponents.

Although segmentation, fractionation, and simplification are all procedures for part-task training, there are important differences among them. In segmentation, the task is decomposed into its constituent elements; however, these elements need not be performed simultaneously, so their precise reintegration is not

critical. In fractionation, concurrent tasks are decomposed into constituent components and trained separately. Careful reintegration is essential for concurrent tasks because critical interrelations between components may only appear when the components are performed simultaneously. Simplification is similar to segmentation in that components of the whole task are trained separately. However, simplification renders the task easier for training purposes because the characteristics of the task are altered, whereas segmentation does not alter the composition of task elements. Of these three procedures, simplification has the greatest relevance to the present investigation.

Simplification was employed in the present investigation. The key to this method is that task components are not only trained individually, but are made less complex to facilitate learning. Simplification is most effective for training high performance skills that are initially very difficult to acquire. Although there is evidence that simplification may not necessarily be more effective overall than whole-task training, it is often cheaper and may be less frustrating than whole-task training for trainees attempting to master an extremely difficult task. Thus, the greatest benefit of simplification accrues for tasks which are highly complex. Simplification need not make the exact, to-be-trained task easier, but may instead utilize training on a highly similar, but easier task.

In the present experiment, three variations of a part-task training component were designed to train declarative knowledge necessary for whole-task performance. Four groups of participants were trained. One group received whole-task practice on the dispatching task throughout the experiment, providing an index of comparison against the part-task conditions. The other three groups received training on a memory-search task consisting of declarative knowledge

elements essential to the performance of the dispatching task. To investigate the effects of providing contextually relevant instructions on training performance, two of the part-task groups were given instructions on exactly how the material being learned would be applied in the whole task. Of these two groups, one actually performed the dispatching task during training, thereby receiving a mixture of part- and whole-task training. The latter condition allowed examination of whether the augmentation of whole-task training with practice on a declarative, part-task component would facilitate task performance. The third part-task group did not receive contextual information about the whole task, and was told only that the declarative information would be used later in a more difficult, complex task. This condition allowed comparison of the effects of part-task training with and without contextually relevant instructions.

Immediately following training, participants in all conditions were transferred to the dispatching task (i.e., whole task). They performed one session of the dispatching task; then the effects of the different part-task procedures were compared to those exhibited in the whole-task condition. To examine the important issue of skill retention based on differential training, participants returned 60 days following transfer for evaluation of differential training on the retention of complex skill.

The present experiment was based on the examination of characteristics of skilled performance in a complex, strategic planning task. Over the past two-and-a-half years, we have developed, tested, and refined what we refer to as the dispatching task. This task was designed to allow the manipulation and examination of important information processing components common to many complex, real-world tasks (e.g., see Fisk et al., 1987; Kyllonen and Woltz,

1989; Salthouse and Somberg, 1982). The information processing components assessed by the dispatching task include visual search, memory scanning, working memory (and the effect of variable memory loads), decision-making, and response selection and execution.

The dispatching task is conceptually similar to tasks performed by a Fighter Duty Officer. The task has several procedural and declarative components, as well as both memory and visual search components. Although the task is conceptually simple, it is quite demanding in its execution. Participants serve as "dispatchers" and, for each trial, receive an order for a specific amount of a cargo to be delivered using a particular vehicle. The dispatcher must select the optimal operator (from a set of four choices) for a given delivery, and must learn rules associated with the determination of load level, load type, and delivery destination characteristics.

Based on each order, the dispatcher's task is to determine the range of possible operators whose licenses qualify them to deliver the cargo. To select the optimal operator with efficiency and accuracy, the dispatcher must learn to associate the names of 27 operators with their corresponding license classifications. Participants have access to extensive help screens via single keystrokes. The help screens provide all the declarative information and rule-based knowledge necessary to perform the task.

The design of the dispatching task used in the present experiment was based on an information-processing, task-analytic methodology developed to isolate trainable skill components across a range of real-world complex tasks. The dispatching task requires memory scanning (participants must hold a list of potential operators in memory); across trials, the number of potential operators (and thus memory load) is manipulated. Participants must also learn rules

associated with performing the task; hence, rule-based learning (necessary for most complex, skill-based tasks) can be assessed. Participants must decide when and how to access help screens most effectively. They must also scan a display to locate the optimal operator.

Method

Subjects. Twenty-four undergraduates (11 males and 13 females) from the Georgia Institute of Technology participated in the experiment. They received a combination of research credit and financial remuneration. Participants were tested for visual acuity of at least 20/40 for near vision and 20/30 for far vision (corrected or uncorrected).

Experimental Task (Dispatching Task). There were two experimental tasks. The first was a dispatching task, presented via a microcomputer, in which participants served as dispatchers for a simulated trucking company. Each trial began with the dispatcher's receipt of an order for a certain amount (in kilograms) of a given cargo to be delivered to a particular destination using a specific vehicle. Participants initiated each trial by pressing the space bar on the microcomputer keyboard. A visual display consisting of the name of the cargo to be delivered, its weight, its destination, and the vehicle to be used for delivery was presented in a two-by-two matrix in tne center of the computer screen.

For each trial, the dispatcher's task was to identify a potential set of vehicle operators qualified to deliver the cargo, based on the requirements of each order. The qualifications of the vehicle operators to make a given delivery corresponded to the level of license they held. When the dispatcher was given a choice of four operators available for each delivery, he or she was required to select the optimal one according to a set of rules. The

fundamental rule was to select the operator with the lowest, or most *minimal license level who was* still qualified to deliver the cargo. Never was there more than a single operator from the same license class in the same trial. In some trials, however, more than one operator was qualified to deliver the cargo. There was always one and only one optimal choice.

Extensive assistance was provided to aid participants in selecting the operators. There were 15 help screens which could be accessed by pressing appropriately labeled keys. Help was divided into three categories: classes, rules, and names. An additional screen which was a "map" of all available help was also provided. Help was available only while the participant was studying information pertaining to the order--never while deciding which operator should deliver the order.

When the participant finished studying the order, he or she pressed the space bar; orientation points (four plus signs arranged in a two-by-two matrix with an "o" centered horizontally and vertically between the plus signs) were then displayed for 500 ms. Immediately following, four operator names were displayed in the same two-by-two matrix. Participants selected an operator by pressing the "7," "9," "1," or "3" key of the numeric keypad. These keys represented the top left, top right, bottom left, and bottom right corners of the two-by-two matrix and were labeled "TL," "TR," "BL," and "BR," respectively.

On incorrect trials, participants received the following feedback. The four operator names remained in the center of the screen, but the correct choice was highlighted using reverse video. At the bottom of the screen, the message "INCORRECT!" was displayed above the message, "The *correct* answer should have been **operator name**, with the optimal choice specified. Finally, the message "You may now

access 'Help.' [illegible] ready, press <+> to exit" was displayed on the top of the screen.

On correct trials, subjects received the following feedback. The four operator names remained in the center of the screen with the correct choice blinking. The word "CORRECT" was displayed inside a box at the top. Directly beneath the box the message "Response Time: XXXX ms" was displayed, where "XXXX" was the actual response latency in milliseconds. Finally, the message "You may now access 'Help.' When ready press <+> to exit" was displayed on the bottom of the screen.

In an earlier study, participants reported that, following commissions of errors, they determined why their responses were wrong by accessing a help screen on the subsequent trial. To ensure that the help accessed prior to responses on any given trial was for the purpose of that trial alone (rather than for feedback on preceding trials), we modified the program to allow participants to access help after selecting the operator within the same trial. Thus, on incorrect trials participants could determine immediately why they had made an error. They could also investigate, if necessary, why they made a correct response.

At the end of each block, participants were given feedback on their mean decision time, percentage of correct responses, and mean total study time (ST). Participants also completed a form on which they recorded this feedback and rated the difficulty of each block on a scale of one to nine--one being easy and nine being hard. This procedure was intended to increase participants' motivation and involvement with the task.

Stimuli (Dispatching Task). The stimuli comprising the basic elements of the experimental task belonged to six classes: (1) cargo, (2) weight, (3) destination, (4)

distance, (5) vehicle, and (6) operator license. The design of the experimental task determined these classes. The metric system (kilograms and kilometers) was used to describe the weights and distances used in the task.

In an attempt to reduce the memory demands of the dispatching task, a set of rules was used to govern construction of the vehicle names. These rules were not provided, but had to be derived by participants. Due to this rule structure, rote memorization of the otherwise arbitrary vehicle names was not necessary.

Construction of the destination and operator license classes involved a different procedure. The operator names associated with each license category and the company names associated with each destination category both were assigned in a wholly arbitrary manner without reliance on an underlying set of logical rules. Thus, learning the operator and company names required the rote memorization of the specific names along with their associated categories. This method is representative of the natural environment, in which names of businesses, truckers, and individuals are selected or assigned in an arbitrary manner.

The names for cargo destinations were derived from the yellow pages of the Atlanta metropolitan area telephone directory. Our principal goal in selecting company names was to minimize any prior associations due to a company or enterprise with which participants might be familiar. Therefore, the chief selection criterion was that the names be nondescript. After selecting a company name from the phone directory, the name was modified by changing its "suffix" to one of the following: Co., Inc., Corp., Ltd., Assoc., Industries, Products, Enterprises, Systems, or Technology. The result was a generic, all-purpose business name (e.g., Ajax, Inc.).

To select the names for human operators to be associated with different license types, the Battig and Montague (1969) category norms were employed. Again, selection criteria were based on an effort to minimize subjects' prior associations or familiarity with operator names (e.g., friends named "Tom"; relatives named "Alice"). First, a list of names that were rated lowest in prototypicality, were a maximum of seven letters in length, and were visually distinct was constructed. Four psychology graduate students, to whom the experimental task was described, were given a list of potential operator names and asked to eliminate any names considered to be unusual, confusable, or unisex.

Because the different subcategories of cargo to be used in the experimental task could be considered natural subcategories (general purpose, liquid, hazardous), we selected cargo names that would be easily, if not naturally, associated with each subcategory. All categories, subcategories, and exemplars are listed within the description of the task presented in Appendix A.

Equipment (Dispatching Task). Epson Equity I+ microcomputers equipped with Epson MBM-2095 monochrome monitors (green phosphor, 50-Hz refresh rate) and Epson multimode graphics adapters were used to present the task. The microcomputers were programmed with Turbo Pascal, version 5.0, to generate files containing task "orders" (see below), present the experimental task, record response behavior, and perform descriptive data analysis. A Heath model AD-1309 white/pink noise generator was used to generate pink noise, which was fed into a Realistic model SA-150 integrated stereo amplifier and output through speakers (approximately 55dB(A)). In this manner external sounds were masked.

Data Collection (Dispatching Task). All keystrokes were captured and stored by the computer program. Hence, a complete record of each subject's use of help was recorded. Also, the time between each keystroke was stored so that it was possible to determine the amount of time spent accessing each screen (e.g., help and study screens). Finally, each subject's decision accuracy (accuracy in choosing the optimal operator in the decision screen) and decision latency on each trial were recorded.

Part-Task Training (Memory-Search Task). The second experimental task was a memory-search task. This task, similar to those typically employed in visual learning/category search paradigms, trained participants on two declarative knowledge components of the dispatcher task. Memory-set size was varied (with either one, two, or three category labels) and display-set size was constant at one exemplar (see the following section on stimuli for more details). There were ten blocks per session and 54 trials per block for a total of 540 trials. Half the trials were positive (target present) and half were negative (target absent).

Each trial proceeded as follows. The memory set was displayed in the left center of the video screen at the beginning of each trial. Participants could study the memory set for up to 30 seconds. To view the display set, participants pressed the space bar on the keyboard. Once the space bar was pressed, an orientation display (a single plus sign) was presented in the same location as the display set for 500 ms. This allowed the participant to focus his/her gaze. Then the display set (either one target exemplar or one distractor exemplar) was presented. The participant's task was to decide as quickly as possible whether a target was present and press a key ("Y" for target

present or "N" for target absent) corresponding to his or her decision.

After each trial and block, participants received performance feedback. After each correct trial, the message, "CORRECT, you responded in X seconds" (where X was their RT in seconds) was displayed. After each false alarm ("Y" selected but "N" correct), the message "ERROR, there was no target present" was displayed. After each miss ("N" selected but "Y" correct), the message, "ERROR, y is a member of z" (where y was the exemplar that had been displayed and z was its superordinate category) was presented. Following each block of trials, the percentage of correct responses and average RT was displayed. Participants were instructed to concentrate on accuracy; if a participant's accuracy fell below 90 percent correct for any block, at the end of that block the program instructed him/her to respond more carefully.

Stimuli (Memory-Search Task). The 18 acronyms for the vehicle operator and destination name classes used in the dispatching task were employed in the memory-search task as memory-set items. The three names associated with each class were used for display-set items. These names and acronyms are contained in Appendix A. Presentation of operator and destination names was alternated across blocks and order was counterbalanced across participants.

Equipment (Memory-Search Task). The equipment was essentially identical to that described previously. Microcomputers, however, were programmed with Psychological Software Tools' Microcomputer Experimental Language (Schneider, 1988) to present and time stimulus displays, record response behavior, and perform descriptive data analysis. Also, an IBM PS/2 Model 30/286 microcomputer with a monochrome VGA monitor was used with one participant.

Subjective Workload Assessment. Subjects in all conditions completed the microcomputer version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) subjective workload scale (Hart and Staveland, 1988) at the end of each session. The NASA-TLX assesses subjective workload on three dimensions relating to task demands (Mental Demand, Physical Demand, and Temporal Demand) and three dimensions relating to the interaction of the subject with the task (Effort, Frustration, and Performance). The six workload dimensions are rated using six bipolar scales. Each scale is represented as a continuous line divided into 20 equally spaced intervals. The scales are anchored by the adjectives "Low" and "High" for the Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration scales. The adjectives "Good" and "Poor" anchor the Performance scale. The NASA-TLX yields scores ranging between 0 and 100. Ratings at the low end of the scale are indicative of low workload and good performance. Ratings at the high end of the scale indicate high workload or poor performance.

After completing the workload scales, participants judged the contribution of each dimension to overall workload in a paired-comparison format. Each dimension was paired sequentially with every other dimension, and the resulting pairs were presented one at a time on the microcomputer screen. Participants chose which of the two dimensions in each pair was more important to the task. A weight for each dimension was derived from these ratings. For each participant, an overall workload score was calculated by multiplying his/her rating of each dimension by its weight and summing the weighted ratings.

Procedure and Design (Dispatching and Memory-Search Tasks). The procedure and design are summarized in

Table 44. Participants were pseudo-randomly assigned to four experimental conditions (six participants per condition): Instruction First, Instruction Last, Alternating, and Whole Task. Assignment was pseudo-random because we sought to place an equal number of males and females in each group; however, the Instruction First group consisted of four females and two males. Each group received five sessions of training.

Whole Task Condition. As its name implies, participants in the Whole Task condition performed the entire dispatching task, but not the memory-search task. The procedure for the training sessions was as follows. Upon participants' arrival on the first day, an experimenter demonstrated the dispatching task by performing three trials. Next, participants were given extensive written instructions for performing the task. These instructions are included in Appendix A. After reading the instructions, participants performed three practice trials; they were permitted to refer to the written instructions while practicing. At the end of the first day, the experimenter gave participants instructions on the NASA-TLX (Hart and Staveland, 1988) then collected subjective workload assessments.

The dispatching task consisted of discrete trials, blocks, and sessions. Participants in the Whole-Task condition performed a total of three training sessions (second through fourth day). There were three 36-trial blocks in each session, for a total of 432 trials. As described previously, each trial represented an "order" to be delivered; the participant's task was to select the optimal operator to deliver that order. Each operator name appeared as the optimal choice an equal number of times within a session. As described previously, a software program generated the files containing these orders.

Condition	Day 1	Day 2	Day 3	Day 4	TRANSFER (Day 5)	RETENTION (Day 65)
Instruction First	Memory-Search Instructions and Practice. Dispatching-Task Instructions and Practice. NASA-TLX	Perform Session 1 Memory-Search Task. NASA-TLX	Perform Session 2 Memory-Search Task. NASA-TLX	Perform Session 3 Memory-Search Task. NASA-TLX	Review Dispatching Task Instructions. Perform Session 4 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.	Review Dispatching Task Instructions. Perform Session 5 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.
Instruction Last	Memory-Search Instructions and Practice. Perform Session 1 Memory-Search Task. NASA-TLX	Perform Session 2 Memory-Search Task. NASA-TLX	Perform Session 3 Memory-Search Task. NASA-TLX	Dispatching Task Instructions and Practice. NASA-TLX	Review Dispatching Task Instructions. Perform Session 4 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.	Review Dispatching Task Instructions. Perform Session 5 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.
Alternating	Memory-Search Instructions and Practice. Dispatching-Task Instructions and Practice. NASA-TLX	Perform Session 1 Memory-Search Task. NASA-TLX	Perform Session 2 Dispatching Task. NASA-TLX	Perform Session 3 Memory-Search Task. NASA-TLX	Review Dispatching Task Instructions. Perform Session 4 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.	Review Dispatching Task Instructions. Perform Session 5 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.
Whole Task	Dispatching-Task Instructions and Practice. NASA-TLX	Perform Session 1 Dispatching Task. NASA-TLX	Perform Session 2 Dispatching Task. NASA-TLX	Perform Session 3 Dispatching Task. NASA-TLX	Review Dispatching Task Instructions. Perform Session 4 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.	Review Dispatching Task Instructions. Perform Session 5 Dispatching Task. NASA-TLX Obtain Comments. Debrief Subjects.

Table 44. Procedure and Design: Section VI

Although presentation of trials was randomly permuted, all participants received the same presentation order.

Participants were also instructed to record, using pen and paper, any comments or observations which they might have regarding the dispatching task. These comments are included in Appendix B. Also, subjects were asked to periodically record their strategies for performing the task. Each day after completing the dispatching task, participants completed the NASA-TLX.

The fifth day consisted of one transfer session. As may be seen in Table 44, the procedure for the transfer session was identical for all participants across all conditions (the other three conditions will be discussed subsequently in greater detail). All participants were given written instructions for performing the dispatching task, which they could review as long as necessary. Next, they performed three identical blocks of the dispatching task then completed the NASA-TLX.

Following completion of the NASA-TLX, participants provided written feedback in response to a series of questions regarding their performance and the task itself. They were then asked to generate the exemplar and class names of the operators and destinations. This provided another measure of the effects of the differential training participants had received. Finally, participants were debriefed and reminded of the retention phase of the experiment.

Following 60 days without practice, participants returned for one session. This retention session was identical to the transfer session. The same set of stimuli were employed and presented in the same order as during the transfer session.

Instruction First. On the first day of training, the Instruction First group first received instructions on performing the memory-search task, then completed two blocks of practice (three trials per block). A study aid consisting of the operator, destination classes, and names was provided. The study aid was constructed by printing the operator and destination names help screens used in the dispatching task (see Appendix A) and laminating them in plastic. Next, an experimenter demonstrated the dispatching task. Then participants were given extensive written instructions for performing the task. After reading the instructions, participants performed three practice trials on the dispatching task. The experimenter explained that they would perform three sessions of the memory-search task and, in the fourth session, perform the dispatching task. Finally, the experimenter instructed participants on the microcomputer version of the NASA-TLX and collected their subjective workload assessments.

On the second, third, and fourth days, the Instruction First group performed Sessions 1, 2, and 3 of the memory-search task. During Session 1 and the first two blocks of Session 2, participants were allowed to use the study aid. They were told that, following the second block of Session 2, they would have to perform the task without the study aid. The experimenter encouraged participants to try to perform the task without the study aid whenever possible. Once the study aid was removed, participants were allowed to review the aid between blocks but not during trials. Each day, upon completion of the memory-search task, they completed the NASA-TLX. On the fifth day, participants were transferred to the dispatching task. The procedure used during the transfer session was discussed in the section describing the Whole Task condition. These participants also returned 60 days later for the retention session described previously.

Instruction Last. The first day of training for participants in the Instruction Last condition was very similar to that for participants in the Instruction First condition. The only difference was that participants in the Instruction Last condition received neither instruction nor practice on the dispatching task. Instead, these participants performed ten blocks of the memory-search task (Session 1). In addition, participants in the Instruction Last condition were told that they would receive instructions and practice on the dispatching task on the fourth day, and perform that task on the fifth day. An experimenter explained that the names used in the memory-search task were the same ones used in the dispatching task; therefore, it was important that they master the names.

During Session 1 and the first two blocks of Session 2, participants in the Instruction Last condition were allowed to use the study aid while performing the trials, but were encouraged to try to perform without it. During Days 2 and 3, they performed Sessions 2 and 3 of the memory-search task. Participants were allowed to review the study aid between blocks but not during trials. At the end of each session they completed the NASA-TLX.

On Day 4, participants received instructions and practice on the dispatching task. First, an experimenter demonstrated the dispatching task by performing three trials. Then participants were given extensive written instructions for performing the task. After reading the instructions, participants performed three practice trials on the dispatching task. Finally, they completed the NASA-TLX. The fifth day was a transfer session in which participants were transferred to the dispatching task. The procedure used during transfer is described in the Whole-Task condition section. These participants also returned 60 days later for the retention session described previously.

Alternating Condition. The first and second days of training for participants in the Alternating condition were identical to those of participants in the Instruction First group. On the third day, however, participants in the Alternating group performed three blocks of the dispatching task (Session 2). After performing the dispatching task, they completed the NASA-TLX.

On the fourth day, participants in the Alternating group performed ten blocks of the memory-search task (Session 3), followed by the NASA-TLX. On the fifth day, participants transferred to the dispatching task. The procedure used during transfer is discussed in the Whole-Task condition section. These participants also returned 60 days later for the retention session described previously.

RESULTS

Memory Search Task Training. Reports by participants and examination of the data indicated that it was more difficult to associate destination names with their respective classification acronyms than to associate operator names with their respective classification acronyms. This led us to analyze separately the blocks in which destinations were used and those in which operators were used (five blocks of each per session). Means and standard deviations of RT, ST, and accuracy (proportion correct) are presented in Tables 45 through 50. Also, plots of these means are presented in Figures 29 through 31.

Participants in the Alternating condition did not perform Session 2 of the memory-search task and used the study aid in Blocks 1 and 2 of Session 3; therefore data were divided into early- and late-training blocks. The early-training blocks consisted of the ten blocks (five of destinations and five of operators) from Session 1 and the

Table 45. Mean Study Time (ms) in Memory Search: Operator Names

Session	Block	Instruction First	Instruction Last	Alternating
1	1	9164.10 (4529.45)	5970.40 (3431.25)	7434.23 (3591.71)
1	2	8614.73 (5496.97)	4813.24 (2826.82)	6322.06 (3508.42)
1	3	7979.74 (4712.87)	4306.18 (2428.88)	6135.68 (3391.81)
1	4	7217.12 (4022.97)	3664.38 (2333.58)	5939.94 (3444.05)
1	5	6877.29 (3904.87)	3536.70 (2175.20)	5024.54 (2949.45)
2	1	6029.70 (3459.81)	4929.29 (3320.47)	
2	2	6664.79 (3640.36)	5955.60 (3731.59)	
2	3	5715.71 (2983.26)	5510.48 (3504.82)	
2	4	5100.44 (2840.55)	5062.05 (3564.68)	
2	5	4681.58 (2375.46)	4492.17 (3474.13)	
3	1	5783.86 (3692.24)	4079.75 (2489.53)	5567.17 (2931.72)
3	2	5071.13 (3177.42)	4846.13 (3468.58)	5541.25 (3030.80)
3	3	4704.74 (3031.58)	3908.53 (2824.33)	4759.87 (2660.47)
3	4	4189.65 (3059.93)	3598.94 (2551.71)	5395.97 (3667.85)
3	5	3683.54 (2693.49)	3441.62 (2724.03)	4374.62 (3246.57)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 46. Mean Reaction Time (ms) in Memory Search: Operator Names

Session	Block	Instruction First	Instruction Last	Alternating
1	1	843.60 (288.38)	1278.49 (807.45)	788.34 (272.00)
1	2	687.38 (190.96)	1149.53 (486.67)	646.02 (117.41)
1	3	675.44 (194.25)	1096.84 (423.79)	627.18 (134.26)
1	4	664.10 (187.65)	1050.05 (419.04)	601.00 (89.32)
1	5	628.02 (164.91)	1012.45 (377.20)	582.48 (76.37)
2	1	620.67 (153.31)	887.59 (330.43)	
2	2	668.10 (188.52)	960.98 (293.13)	
2	3	669.49 (158.18)	959.37 (293.61)	
2	4	666.30 (167.57)	987.08 (335.98)	
2	5	617.58 (109.54)	883.26 (242.98)	
3	1	661.29 (195.79)	909.59 (326.85)	586.49 (104.45)
3	2	672.22 (155.41)	818.94 (244.70)	685.43 (138.38)
3	3	664.21 (167.05)	909.35 (329.63)	653.33 (112.59)
3	4	668.09 (139.10)	874.63 (317.81)	682.59 (174.76)
3	5	685.40 (194.32)	842.89 (281.50)	750.45 (224.68)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 47. Mean Proportion Correct in Memory Search: Operator Names

Session	Block	Instruction First	Instruction Last	Alternating
1	1	.91 (.10)	.88 (.17)	.92 (.10)
1	2	.95 (.06)	.94 (.08)	.95 (.08)
1	3	.95 (.07)	.95 (.07)	.91 (.09)
1	4	.95 (.08)	.95 (.08)	.92 (.12)
1	5	.95 (.10)	.94 (.08)	.92 (.13)
2	1	.92 (.12)	.95 (.07)	
2	2	.90 (.15)	.76 (.24)	
2	3	.92 (.12)	.86 (.11)	
2	4	.93 (.11)	.90 (.09)	
2	5	.92 (.11)	.91 (.10)	
3	1	.91 (.12)	.87 (.15)	.92 (.09)
3	2	.89 (.14)	.91 (.10)	.88 (.12)
3	3	.95 (.08)	.97 (.12)	.87 (.14)
3	4	.96 (.07)	.96 (.06)	.91 (.09)
3	5	.95 (.08)	.96 (.06)	.92 (.11)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 48. Mean Study Time (ms) in Memory Search: Destination Names

Session	Block	Instruction First	Instruction Last	Alternating
1	1	8300.22 (6020.41)	6535.13 (3682.16)	9442.51 (4809.73)
1	2	8899.43 (5289.33)	6049.07 (3749.88)	7808.81 (4382.57)
1	3	8784.21 (5264.95)	5479.76 (3803.71)	7525.01 (4282.99)
1	4	8407.83 (4568.21)	5174.80 (3249.84)	6701.94 (3643.25)
1	5	7821.22 (4172.31)	4957.16 (3390.76)	5883.24 (3450.25)
2	1	6207.86 (3073.23)	6416.36 (4541.43)	
2	2	6975.18 (3378.41)	7728.67 (4426.68)	
2	3	6398.20 (3366.35)	5875.02 (3613.15)	
2	4	5712.99 (2748.93)	5661.66 (3321.06)	
2	5	5255.17 (2640.54)	4391.13 (2369.41)	
3	1	5708.66 (2922.54)	5150.36 (2874.99)	6358.50 (3635.48)
3	2	5890.40 (2845.67)	5051.95 (3564.72)	5147.40 (3254.21)
3	3	5276.66 (3022.80)	4489.46 (3002.89)	5543.61 (3655.98)
3	4	4396.64 (2386.39)	4199.09 (2853.18)	5630.87 (3859.40)
3	5	4438.06 (2649.47)	4458.62 (2945.78)	5669.60 (4073.37)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 49. Mean Reaction Time (ms) in Memory Search: Destination Names

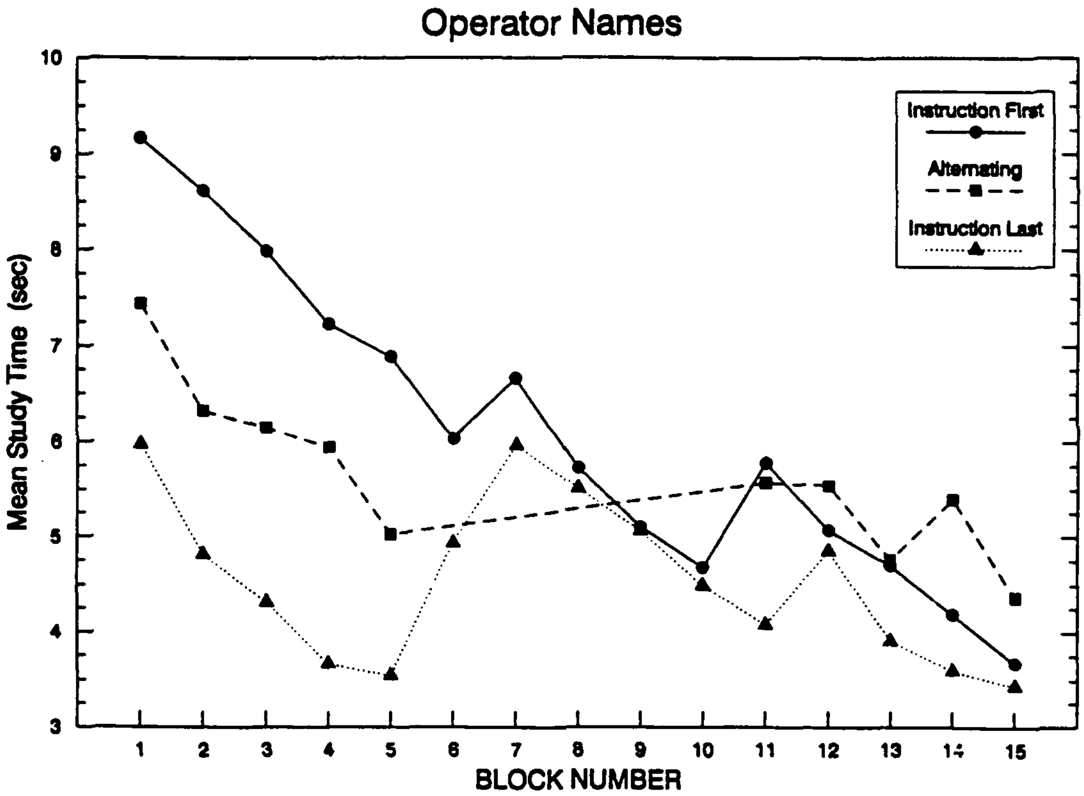
Session	Block	Instruction First	Instruction Last	Alternating
1	1	857.96 (280.34)	1454.30 (680.97)	837.46 (382.73)
1	2	676.25 (183.69)	1325.54 (587.68)	666.65 (141.77)
1	3	700.00 (197.29)	1175.03 (444.37)	643.00 (127.08)
1	4	666.93 (183.03)	1166.97 (469.61)	656.20 (129.82)
1	5	654.47 (233.99)	1133.06 (504.96)	676.00 (161.27)
2	1	627.15 (145.21)	918.72 (333.72)	
2	2	707.91 (194.12)	949.39 (277.52)	
2	3	725.51 (193.31)	1021.68 (380.99)	
2	4	724.68 (179.13)	965.30 (259.21)	
2	5	684.95 (174.55)	939.16 (214.72)	
3	1	643.91 (162.90)	878.75 (172.22)	670.91 (237.18)
3	2	656.02 (159.05)	956.46 (283.46)	695.68 (170.63)
3	3	716.32 (244.85)	875.25 (250.31)	699.41 (150.78)
3	4	698.19 (202.62)	860.08 (210.69)	726.89 (201.15)
3	5	712.82 (214.17)	882.79 (217.37)	700.47 (158.83)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 50. Mean Proportion Correct in Memory Search: Destination Names

Session	Block	Instruction First	Instruction Last	Alternating
1	1	.93 (.10)	.90 (.17)	.94 (.08)
1	2	.93 (.11)	.91 (.11)	.94 (.10)
1	3	.94 (.08)	.96 (.07)	.93 (.11)
1	4	.92 (.10)	.93 (.08)	.90 (.12)
1	5	.94 (.09)	.95 (.07)	.91 (.13)
2	1	.92 (.11)	.92 (.09)	
2	2	.88 (.17)	.73 (.23)	
2	3	.88 (.14)	.80 (.17)	
2	4	.89 (.15)	.81 (.16)	
2	5	.89 (.18)	.85 (.13)	
3	1	.90 (.13)	.83 (.16)	.92 (.11)
3	2	.93 (.09)	.88 (.12)	.87 (.15)
3	3	.93 (.09)	.93 (.07)	.89 (.12)
3	4	.91 (.13)	.94 (.09)	.90 (.11)
3	5	.95 (.09)	.95 (.07)	.89 (.13)

NOTE: Means are on top; standard deviations are in parentheses underneath.



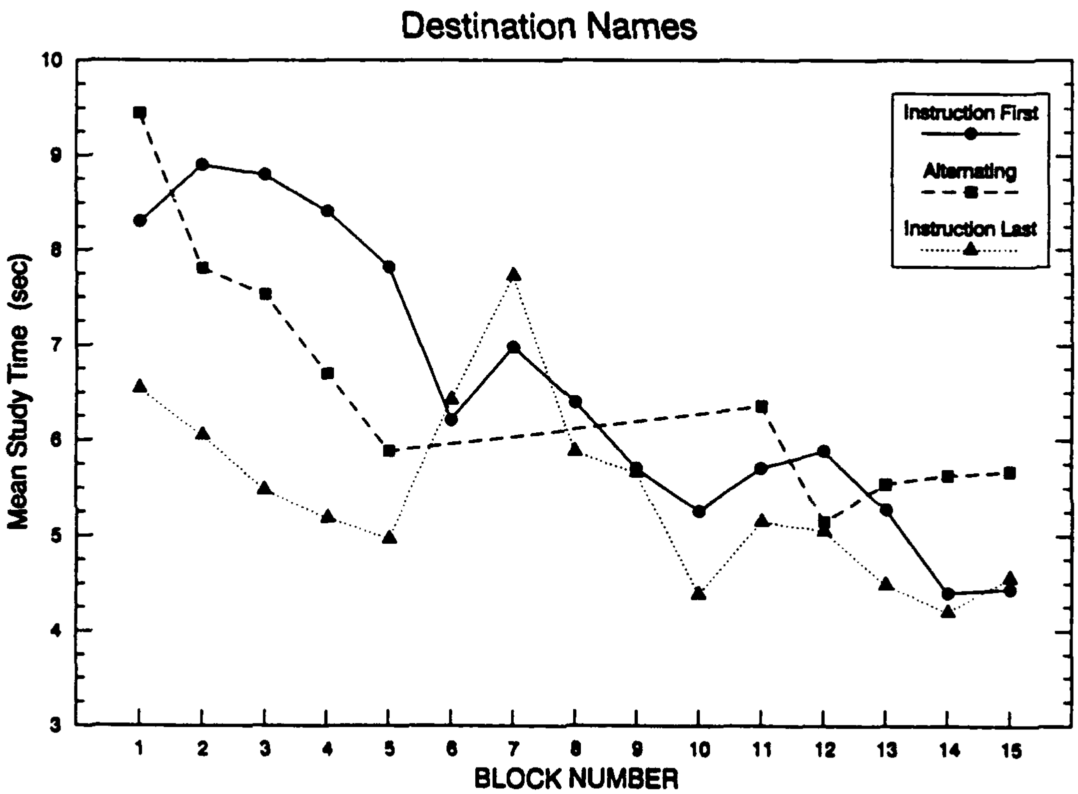
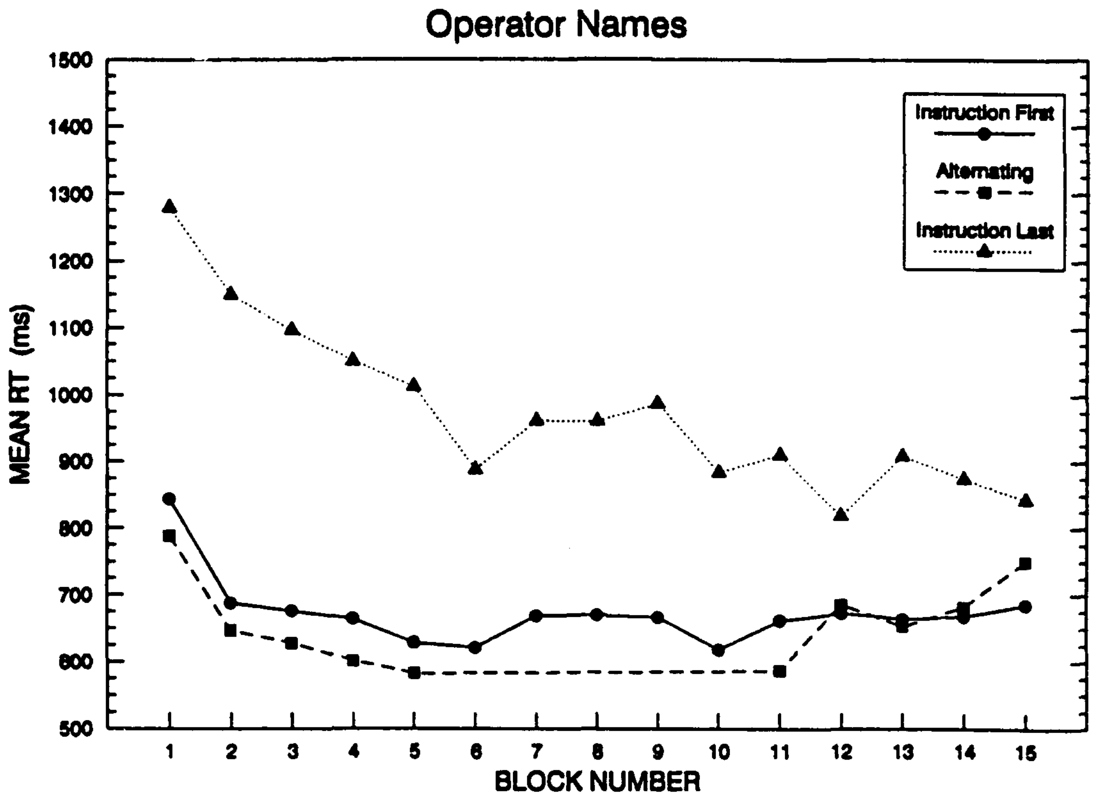


Figure 29. Mean Study Time as a Function of Block



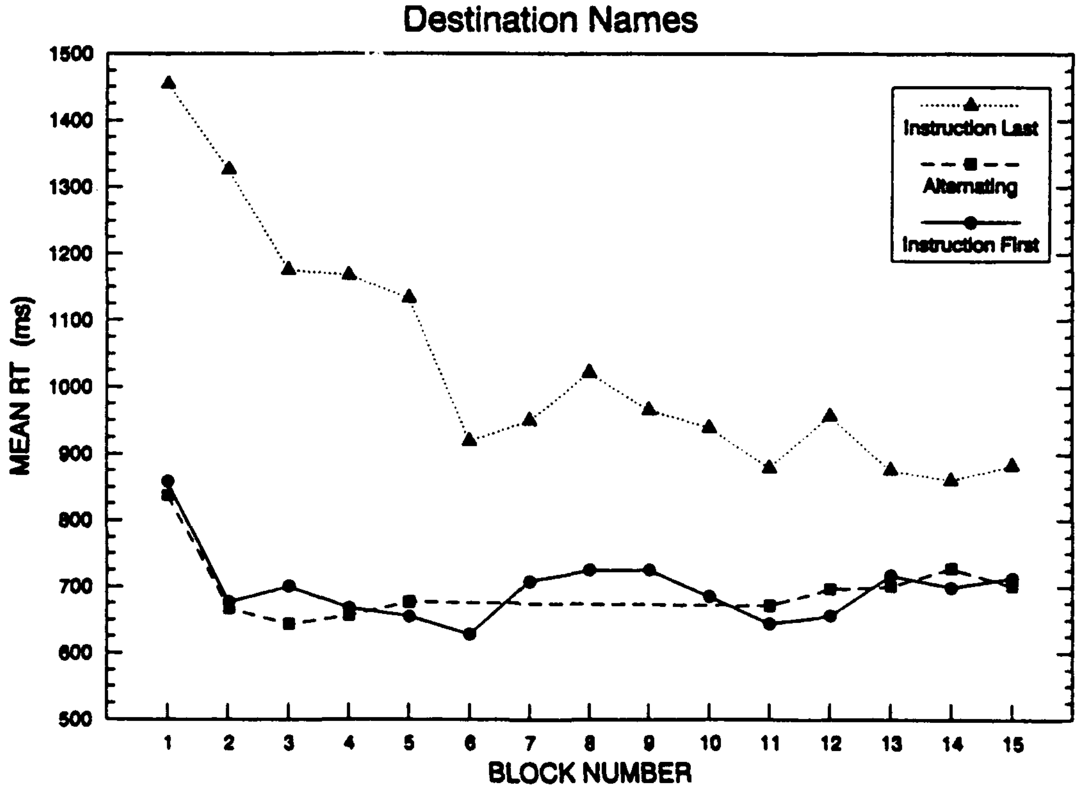


Figure 30. Mean Reaction Time as a Function of Block

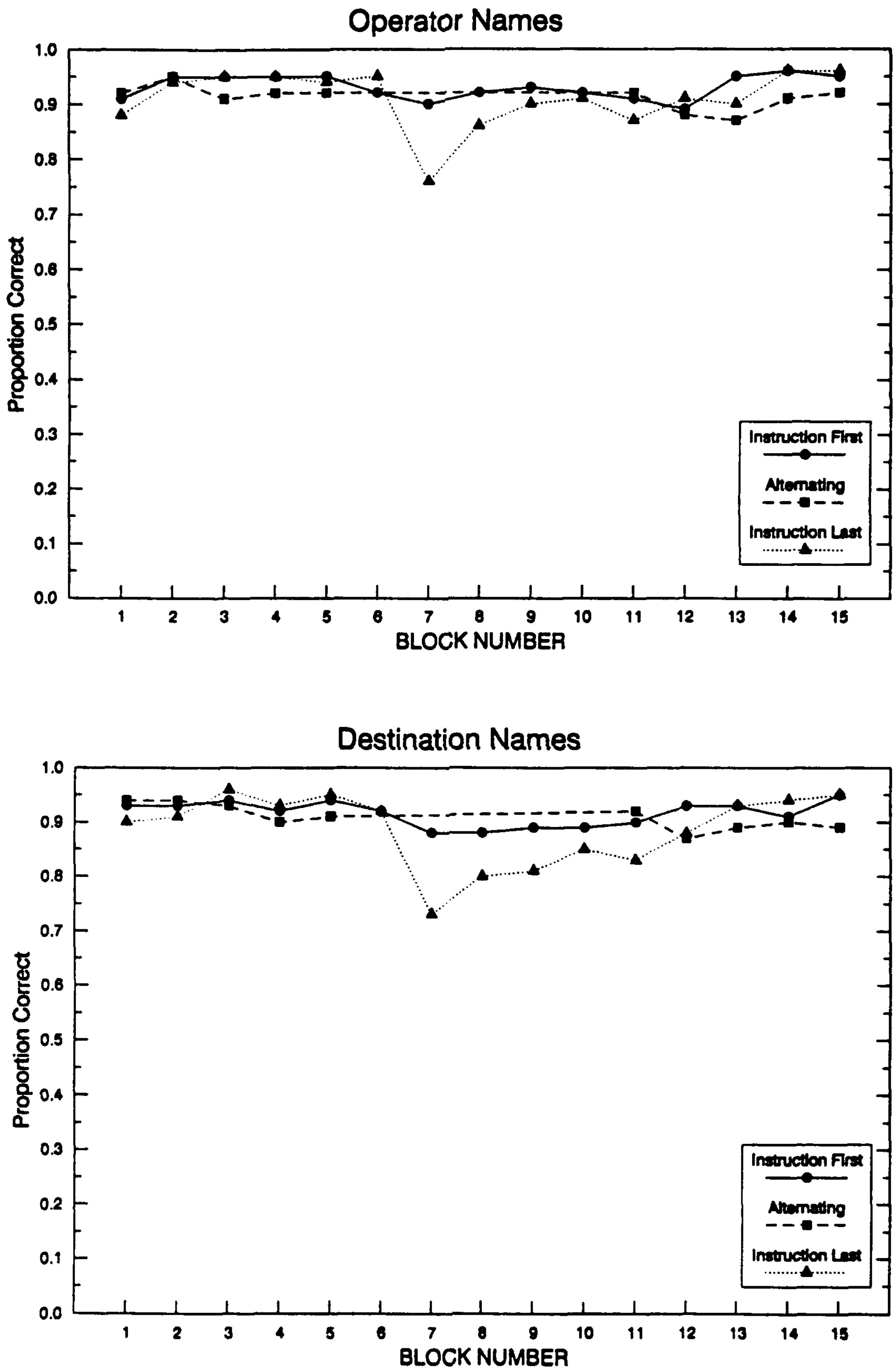


Figure 31 . Mean Proportion Correct as a Function of Block

late-training blocks consisted of the final eight blocks (four of destinations and four of operators) from Session 3.

Operator Names. A 3 x 5 factorial ANOVA was performed analyzing ST as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) for trials using operator names early in training. A summary of this analysis is reported in Table 51; relevant mean scores are reported in Table 45. The main effects for condition ($F(2, 15) = 10.82$, $p < .01$) and block ($F(4, 60) = 10.81$, $p < .01$) were significant. Figure 29 demonstrates this result, revealing a dramatic decline in ST from the first block to the fifth block across all three conditions. Also, a more than one-second mean difference between each of the three conditions persisted across the five blocks.

ST for trials using operator names late in training was also analyzed as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3) in a 3 x 4 factorial ANOVA. This analysis is summarized in Table 51, and the relevant mean scores are reported in Table 45. The only statistically significant effect was a main effect for block ($F(3, 45) = 6.04$, $p < .01$). This result indicates that participants continued to improve after use of the study guide was eliminated. Also, participants in all three conditions required about the same amount of time to study the classification acronyms, although those in the Alternating condition appear to be somewhat slower (probably due to less practice at the task).

RT as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) was analyzed in a 3 x 5 factorial ANOVA for trials using operator names early in training. A summary of this analysis is reported in

Table 51. Summary of Analysis of Variances for Memory-Search Task, Operator Names Trials: Study Time

Early in Training

Source	df	MS	F
condition	2	92546.64	10.82**
subjects w/in condition	15	8552.37	
block	4	15131.62	10.81**
block x condition	8	313.12	0.22
block by subjects w/in condition	60	1399.30	

Late in Training

Source	df	MS	F
condition	2	6898.57	0.55
subjects w/in condition	15	12626.37	
block	3	5262.43	6.04**
block x condition	6	552.70	0.63
block by subjects w/in condition	45	871.18	

** $p < .01$

Table 52; relevant mean scores are reported in Table 46. Main effects of condition ($F(2, 15) = 8.00$, $p < .01$) and block ($F(4, 60) = 6.08$, $p < .01$) were significant. Figure 30 shows that participants in the Instruction Last condition were markedly slower than participants in either of the other two conditions. RTs declined more than 200 ms from Block 1 to Block 5 for participants in the Instruction First and Alternating conditions, and more than 250 ms for participants in the Instruction Last condition.

A 3 x 4 factorial ANOVA was performed analyzing RT as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3) for trials using operator names late in training. This analysis is summarized in Table 52; relevant mean scores are reported in Table 46. There were no statistically significant effects; however, Figure 30 shows that, across the final four training blocks, participants in the Instruction Last condition were consistently slower than participants in the other two conditions. It is likely that this difference was not statistically significant because of the small number of participants in each condition. It is also worth noting that RT performance in the Instruction First and Last conditions shows little change across the final four blocks, while performance in the Alternating condition actually declines by over 150 ms. Again, these participants had one session less of practice.

A 3 x 5 factorial ANOVA was performed analyzing the dependent variable of proportion correct as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) for trials involving operator names early in training. This analysis is summarized in Table 53; relevant mean scores are

Table 52. Summary of Analysis of Variances for Memory-Search Task, Operator Names Trials: Reaction Time

Early in Training

Source	df	MS	F
condition	2	1982.79	8.00**
subjects w/in condition	15	247.92	
block	4	142.03	6.08**
block x condition	8	2.06	0.09
block by subjects w/in condition	60	23.37	

==

Late in Training

Source	df	MS	F
condition	2	258.08	3.01
subjects w/in condition	15	85.62	
block	3	3.48	0.81
block x condition	6	8.17	1.90
block by subjects w/in condition	45	4.31	

** $p < .01$

Table 53. Summary of Analysis of Variances for Memory-Search Task, Operator Names Trials: Proportion Correct (Accuracy)

Early in Training

Source	df	MS	F
condition	2	0.003	0.79
subjects w/in condition	15	0.004	
block	4	0.005	3.11*
block x condition	8	0.002	1.05
block by subjects w/in condition	60	0.002	

Late in Training

Source	df	MS	F
condition	2	0.013	1.73
subjects w/in condition	15	0.007	
block	3	0.012	8.13**
block x condition	6	0.002	1.38
block by subjects w/in condition	45	0.002	

* $p < .05$

** $p < .01$

reported in Table 47. The only statistically significant effect was the main effect for block ($F(4, 60) = 3.11$, $p < .05$). Figure 31 demonstrates an improvement in accuracy for all three conditions from Block 1 to Block 2. Following Block 2, accuracy appears stable across the remaining three blocks.

Proportion correct as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3) for trials involving operator names late in training was analyzed in a 3 x 4 factorial ANOVA. A summary of this analysis is provided in Table 53; relevant mean scores are reported in Table 47. Again, the only statistically significant effect was the main effect for block ($F(3, 45) = 8.13$, $p < .01$). This result indicates that accuracy performance was roughly equivalent across all three conditions at the end of training. As Figure 31 reveals, accuracy actually improved across the remaining blocks.

Destination Names. A 3 x 5 factorial ANOVA was performed on ST as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) for trials involving destination names early in training. A summary of this analysis is reported in Table 54 and the relevant mean scores are reported in Table 48. The main effects for condition ($F(2, 15) = 4.65$, $p < .05$) and block ($F(4, 60) = 3.61$, $p < .05$) were significant. As was the case with operator names, there was a dramatic decline in ST from Block 1 to Block 5 in all three conditions. The same pattern of differential STs was demonstrated across the three conditions: participants in the Instruction First condition spent the most time studying the classification acronyms and destination names and participants in the Instruction Last condition spent the least time studying.

Table 54. Summary of Analysis of Variances for Memory-Search Task, Destination Names Trials: Study Time

Early in Training

Source	df	MS	F
condition	2	60804.07	4.65*
subjects w/in condition	15	13073.50	
block	4	9449.90	3.61*
block x condition	8	2419.71	0.92
block by subjects w/in condition	60	2618.11	

Late in Training

Source	df	MS	F
condition	2	5122.86	0.40
subjects w/in condition	15	12813.71	
block	3	1314.43	1.49
block x condition	6	1440.43	1.64
block by subjects w/in condition	45	879.71	

* $p < .05$

Using ST as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3), a 3 x 4 factorial ANOVA was performed for trials involving destination names late in training. A summary of this analysis is presented in Table 54; relevant mean scores are reported in Table 48. There were no statistically significant effects. However, Figure 29 indicates that STs increased slightly across the final four blocks of Session 3 for participants in the Alternating condition. This is probably due to the fact that participants in the Alternating condition had the study guide removed immediately prior to the final four blocks of Session 3. STs actually appear to be declining in the other two conditions, suggesting a block by condition interaction. However, there was insufficient power to detect this effect.

A 3 x 5 factorial ANOVA was performed on RT as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) for trials using destination names early in training. A summary of this analysis is provided in Table 55; relevant mean scores are reported in Table 49. The main effects for condition ($\underline{F}(2, 15) = 9.48$, $\underline{p} < .01$) and block ($\underline{F}(4, 60) = 11.10$, $\underline{p} < .01$) were significant. As was the case with operator names, RTs in the Instruction Last condition were dramatically slower than those in the other two conditions. However, an impressive decline in RT is demonstrated in all three conditions: decreases of 200, 320, and 160 ms in the Instruction First, Instruction Last, and Alternating conditions, respectively.

A 3 x 4 factorial ANOVA was performed analyzing RT as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3) for trials using destination names late in training. This analysis is summarized in Table 55;

Table 55. Summary of Analysis of Variances for Memory-Search Task, Destination Names Trials: Reaction Time

Early in Training

Source	df	MS	F
condition	2	2999.19	9.48**
subjects w/in condition	15	316.51	
block	4	163.75	11.10**
block x condition	8	13.64	0.93
block by subjects w/in condition	60	14.75	

Late in Training

Source	df	MS	F
condition	2	298.32	4.33*
subjects w/in condition	15	68.96	
block	3	0.19	0.06
block x condition	6	8.35	2.75*
block by subjects w/in condition	45	3.04	

* $p < .05$

** $p < .01$

relevant mean scores are reported in Table 49. Although there was a statistically significant main effect for condition, this was of less interest than the significant interaction of block with condition ($F(6, 45) = 2.75$, $p < .05$). This interaction is plotted in Figure 30. The interaction appears to be due to an approximately 80-ms decline in the Instruction Last condition from Block 2 to Block 3, while RT increased by 60 ms in the Instruction First condition and remained stable in the Alternating condition.

A 3 x 5 factorial ANOVA was performed using proportion correct as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 1 through 5, Session 1) for trials using destination names early in training. A summary of this analysis is reported in Table 56; relevant mean scores are reported in Table 50. There were no statistically significant effects. Accuracies tended to hover around 0.93 proportion correct. This result is not surprising because all participants were using the study aid.

Proportion correct as a function of training condition (Instruction First, Instruction Last, and Alternating) and practice (Blocks 2 through 5, Session 3) was analyzed in a 3 x 4 factorial ANOVA for trials using destination names late in training. This analysis is summarized in Table 56; relevant mean scores are reported in Table 50. There were no statistically significant effects. It may be noted that, although accuracy performance is quite stable in the Instruction First and Alternating conditions, it improved steadily from 0.88 to 0.95 proportion correct in the Instruction Last condition.

Dispatching Task Training. Participants in the Alternating and Whole Task conditions received training on the dispatching task. Those in the Whole Task condition

Table 56. Summary of Analysis of Variances for Memory-Search Task, Destination Names Trials: Proportion Correct (Accuracy)

Early in Training

Source	df	MS	F
condition	2	0.001	0.20
subjects w/in condition	15	0.004	
block	4	0.002	1.23
block x condition	8	0.002	1.46
block by subjects w/in condition	60	0.002	

Late in Training

Source	df	MS	F
condition	2	0.016	1.50
subjects w/in condition	15	0.010	
block	3	0.004	2.38
block x condition	6	0.002	1.20
block by subjects w/in condition	45	0.002	

were trained on the dispatching task during Sessions 1, 2, and 3. Participants in the Alternating condition, however, performed the dispatching task only during training Session 2. A number of performance indices were obtained for the dispatching task. We concentrated on two measures: proportion correct and total time required to reach a decision (total time). Total time consists of the sum of time spent accessing all help screens prior to selecting an operator (pre-response help), studying the order, and selecting an operator (decision latency). Other performance indices (e.g., number of keys pressed, time spent in post-response help, ST, and decision latency) are included in Appendix C.

Mean total time and proportion correct by block and session are presented in Tables 57 and 58. The block means are plotted, respectively, as speed and accuracy in Figure 32. Examination of Whole Task performance reveals a dramatic decline in total time from Block 1 of Session 1 through Block 2 of Session 2, declining from an average of almost 64 s per block to approximately 19 s. Total time performance continued to improve modestly, declining another 5 s by the end of Session 3. Whole Task accuracy improved steadily from 0.75 proportion correct during the first block of Session 1 to 0.94 proportion correct during Session 2, where it leveled off.

A striking improvement is also found in the performance of participants in the Alternating condition in their single session of training on the dispatching task. Initially, total time performance in the Alternating condition was about 6.5 s slower than that of the Whole Task condition; however, by Block 3 of Session 2, Alternating Condition participants were about 2.8 s faster than their more practiced counterparts. In three blocks they reduced total time by 20 s. Accuracy also improved dramatically--from

Table 57. Mean Total Time to Reach Decision (seconds)

Session	Block	Instruct. First	Instruct. Last	Alternat.	Whole Task
1	1				63.91 (46.26)
1	2				35.04 (20.10)
1	3				26.21 (15.69)
2	1			35.73 (30.81)	29.20 (16.37)
2	2			20.55 (10.58)	19.26 (8.25)
2	3			15.92 (7.77)	18.77 (9.79)
3	1				17.16 (9.88)
3	2				15.95 (9.43)
3	3				14.06 (9.40)
Transfer	1	28.13 (21.84)	32.86 (23.15)	13.44 (8.22)	15.30 (9.71)
Transfer	2	17.18 (9.42)	19.98 (10.22)	10.98 (5.36)	13.75 (9.89)
Transfer	3	15.85 (10.50)	17.55 (10.24)	10.45 (4.70)	13.47 (8.85)
Retention	1	28.37 (39.85)	25.61 (18.58)	16.60 (11.16)	27.82 (26.67)
Retention	2	17.08 (14.41)	17.69 (9.25)	13.11 (6.28)	18.61 (12.97)
Retention	3	5.02 (9.21)	15.76 (8.57)	12.26 (5.54)	16.53 (10.26)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 53. Mean Proportion Correct

Session	Block	Instruct. First	Instruct. Last	Alternat.	Whole Task
1	1				.75 (.13)
1	2				.82 (.15)
1	3				.85 (.15)
2	1			.71 (.11)	.87 (.15)
2	2			.79 (.12)	.90 (.10)
2	3			.82 (.09)	.94 (.12)
3	1				.94 (.67)
3	2				.90 (.15)
3	3				.94 (.06)
Transfer	1	.82 (.11)	.76 (.13)	.87 (.08)	.96 (.07)
Transfer	2	.84 (.10)	.82 (.05)	.85 (.12)	.94 (.06)
Transfer	3	.85 (.12)	.80 (.07)	.85 (.11)	.95 (.06)
Retention	1	.75 (.17)	.67 (.11)	.76 (.10)	.92 (.08)
Retention	2	.82 (.22)	.69 (.15)	.80 (.08)	.91 (.08)
Retention	3	.85 (.18)	.77 (.20)	.82 (.08)	.97 (.03)

NOTE: Means are on top; standard deviations are in parentheses underneath.

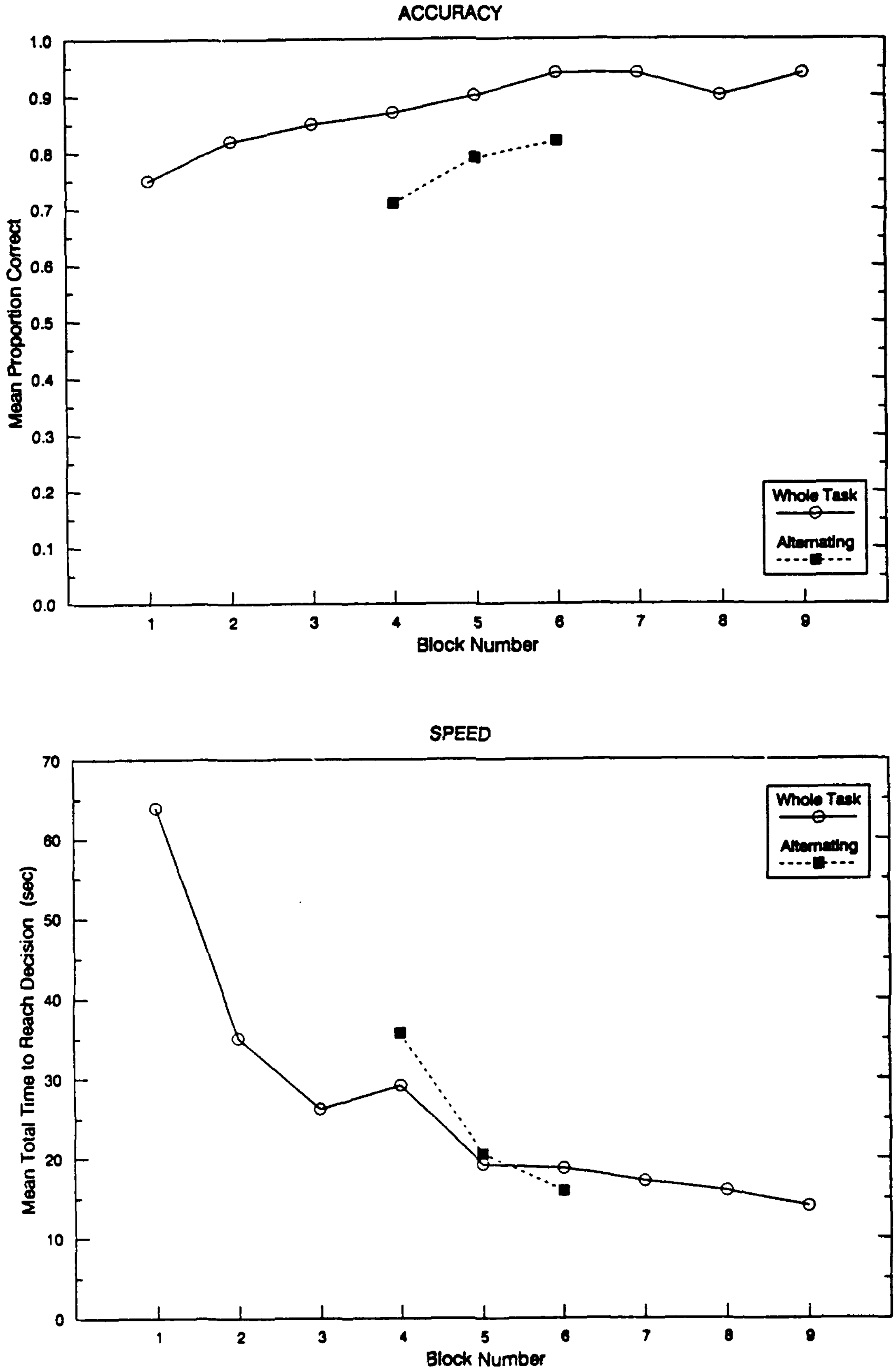


Figure 32. Speed and Accuracy as Functions of Training

0.71 proportion correct in the first block to 0.82 in the final block.

In comparing Alternating condition performance in the first session of the dispatching task with performance in the Whole Task condition (i.e., Alternating in Session 2 with Whole Task in Session 1), it is clear that total time performance in the Alternating condition was vastly superior to that in the Whole Tas condition. In Session 1, total time for particpants in tne Whole Task condition was about 41.7 s. In Session 2, average total time in the Alternating condition was only about 24 s, a difference of more than 17.6 s. Participants in the Whole Task condition, however, were somewhat more accurate in their first session: 0.81, as compared to 0.77 proportion correct for first session performance in the Alternating condition.

<u>Transfer</u>. To compare performance on the dispatching task at transfer in the four training conditions, total time and proportion correct were examined. A 4 x 3 factorial ANOVA was performed analyzing total time as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and block (Blocks 1 through 3, Session 4). A summary of this analysis is presented in Table 59; relevant means are presented in Table 57. Although both main effects were significant, these were of less interest than the finding of a significant two-way interaction (training condition x block, $\underline{F}(6, 40) = 16.97$, $\underline{p} < .01$).

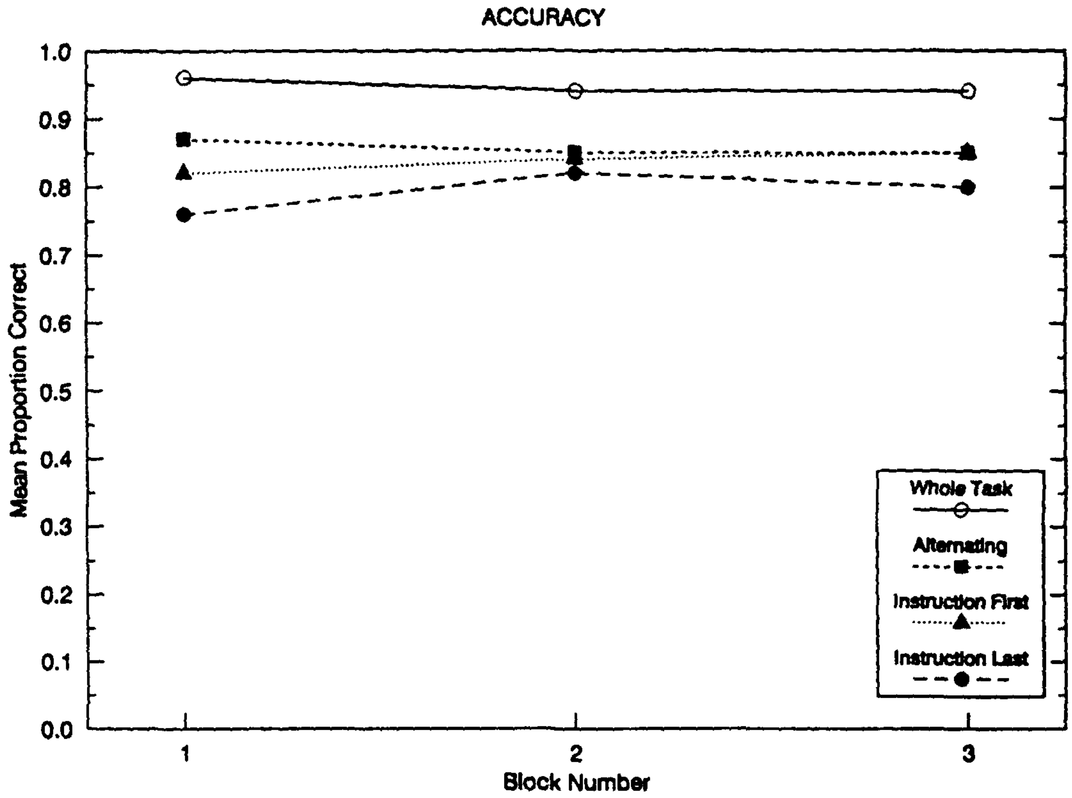
For ease of interpretation, mean total time for each condition was plotted as a function of block and is presented in Figure 33. There was a modest improvement in total time performance from the first to the second block in the two conditions in which participants received at least one session of training on the dispatching task (a reduction in total time of about 2.5 s and 1.5 s in the Alternating

Table 59. Summary of Analysis of Variance for Dispatching Task, Transfer Session: *Total Decision Time*

Transfer

Source	df	MS	F
condition	3	537004.11	6.04**
subjects w/in condition	20	88965.22	
block	2	461514.08	98.77**
block x condition	6	79303.22	16.97**
block by subjects w/in condition	40	4672.79	

** $\underline{p} < .01$



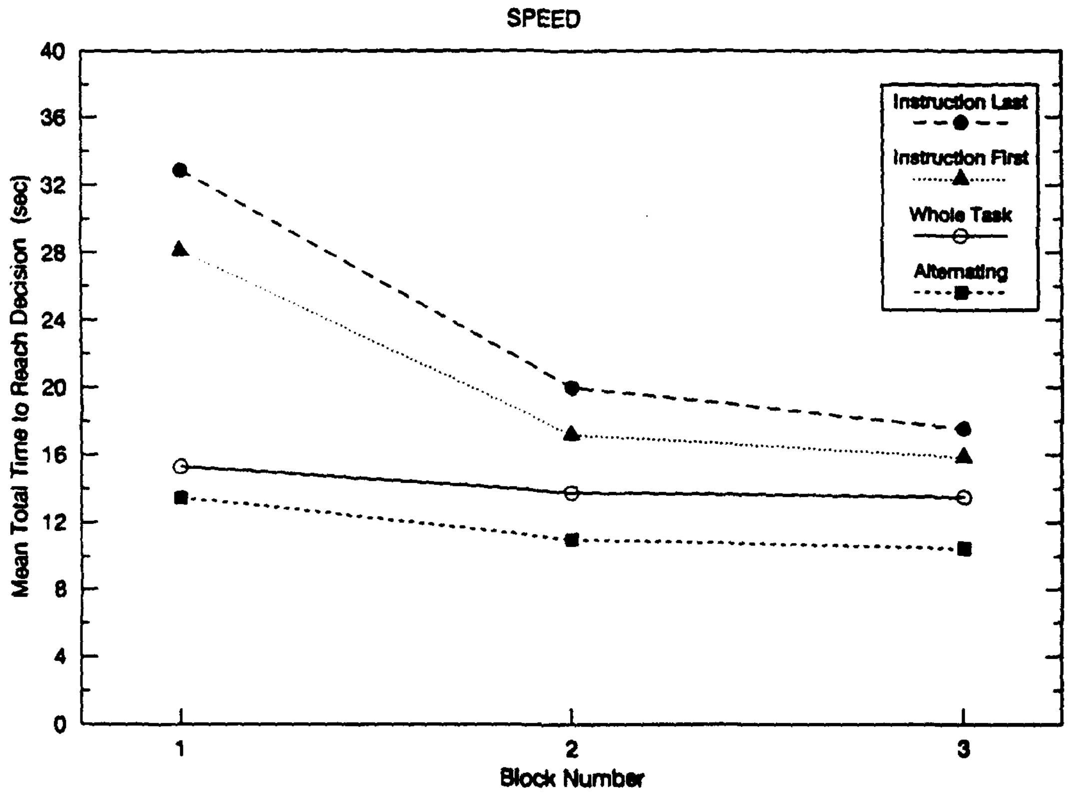


Figure 33. Speed and Accuracy as a Function of Block (Transfer)

and Whole Task conditions, respectively). By contrast, in the two conditions in which participants received training only on the memory-search task, there was marked improvement in total time performance from Block 1 to Block 2 (almost 13 s and 11 s in the Instruction Last and Instruction First conditions, respectively) and modest improvement in the remaining block (over 2 s and 1 s, respectively). This demonstrated pattern of total time performance is the most dramatic effect and is clearly defined by whether or not participants received training on the dispatching task. While significant learning took place in the memory-search-only conditions, more stable performance was evidenced in the two dispatching task conditions. As may be expected, total time performance was superior in the two conditions which provided participants with dispatching-task experience, although it appeared that the four conditions were beginning to converge by the end of the transfer session.

Accuracy performance (proportion correct) as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and block (Blocks 1 through 3, Session 5) were also analyzed in a 4 x 3 factorial ANOVA. This analysis is summarized in Table 60; means for each condition are presented in Table 58. The main effect of condition was the only statistically significant effect ($\underline{F}(3, 20) = .4.66$, $\underline{p} < .01$). In Figure 33, a plot of the mean proportion correct as a function of block for each training condition is presented. As the figure illustrates, accuracy in the Whole Task condition remained at the 0.94 level or higher across all three blocks. Performance in this condition was at least 0.09 points better than the next closest condition, Alternating (although this difference was not statistically significant). There is clearly an ordering of performance at Block 1, with Instruction First lagging 0.05 points behind Alternating; and Instruction

Table 60. Summary of Analysis of Variance for Dispatching Task, Transfer Session: Proportion Correct

Transfer

Source	df	MS	F
condition	3	0.083	4.66**
subjects w/in condition	20	0.018	
block	2	0.001	
block x condition	6	0.002	
block by subjects w/in condition	40	0.004	

** $p < .01$

Last, in turn, lagging 0.06 points behind Alternating. Performance in all three part-task conditions converged at 0.84 proportion correct at Block 2. However, at Block 3, performance in the Instruction Last condition declined to 0.80, while Alternating and Instruction First remained stable at 0.85.

Retention. As was done for the transfer data, speed (total time) and accuracy (proportion correct) performance on the dispatching task during the retention session was compared across the four training conditions. Total time as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and block (Blocks 1 through 3, Session 5) was analyzed in a 4 x 3 factorial ANOVA. Table 61 presents a summary of this analysis; the relevant means are shown in Table 57. Once again, the two-way interaction (training condition x block) was significant ($F(6, 40) = 3.15$, $p < .01$). Figure 34 is a plot of this interaction. Total time performance in the Alternating condition at Block 1 was greatly superior to the next closest condition, Instruction Last, with a difference of 9 s. In every condition, with the exception of Alternating, there was a dramatic improvement in total time performance from the first to the second block. Total time performance in the Whole Task, Instruction First, and Instruction Last conditions was virtually identical across the three retention blocks. Finally, while at Block 1 there was more than 11.75 s difference between the best and worst conditions, by the end of the retention session the three conditions began to converge and this difference diminished to 4.25 s (of course, this is still a sizeable difference).

A 4 x 3 factorial ANOVA was performed on proportion correct as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and

Table C1. Summary of Analysis of Variance for Dispatching Task, Retention Session: Total Decision Time

Retention

Source	df	MS	F
condition	3	182886.32	2.12
subjects w/in condition	20	86431.17	
block	2	643553.04	77.67**
block x condition	6	26068.15	3.15**
block by subjects w/in condition	40	8285.70	

** $p < .01$

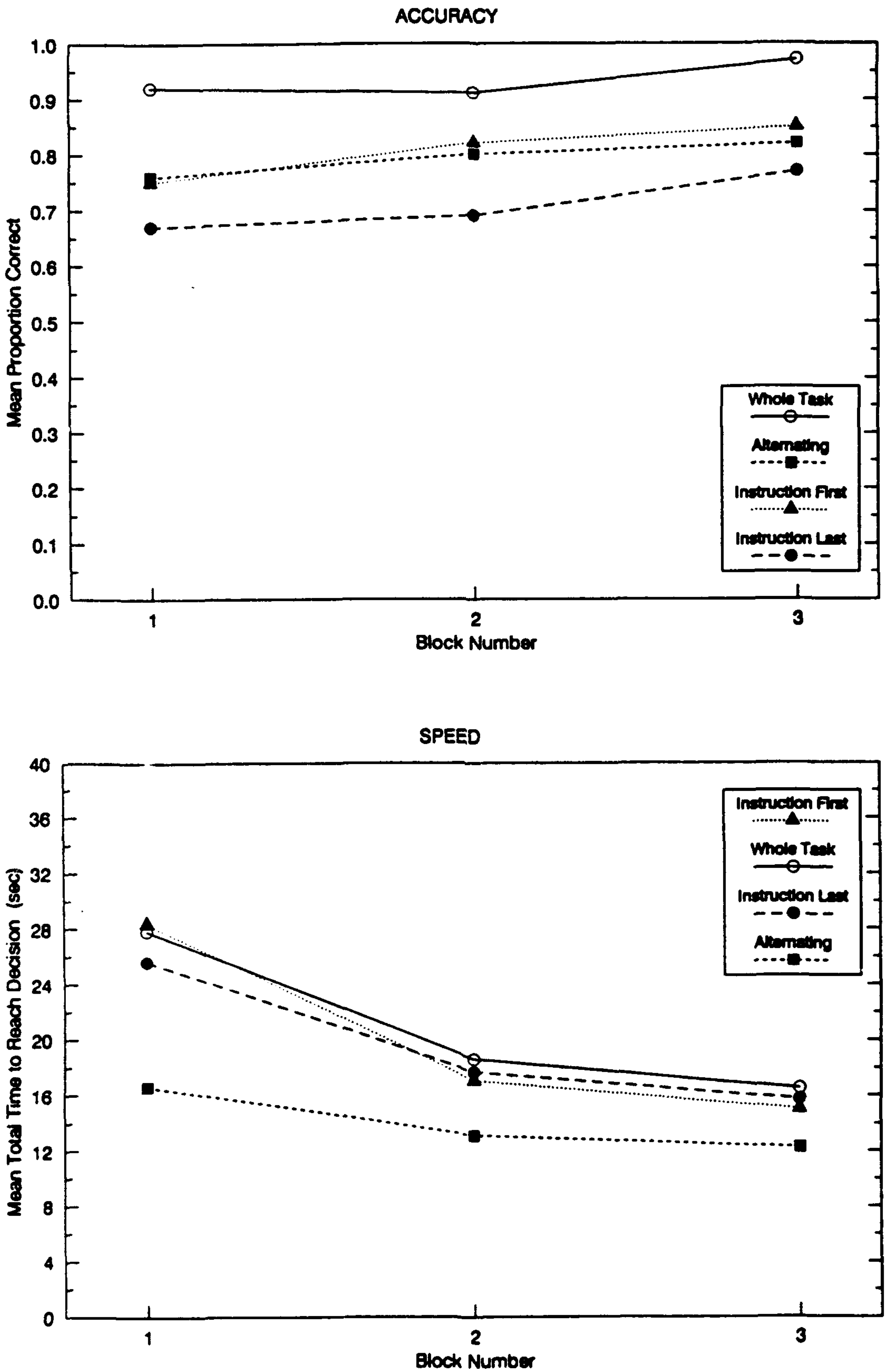


Figure 34. Speed and Accuracy as a Function of Block (Retention)

block (Blocks 1 through 3, Session 5). This analysis is summarized in Table 62; condition means are reported in Table 58. The main effects of both training condition and block were statistically significant ($F(3, 20) = 3.26$, $p < .05$ and $F(2, 40) = 10.41$, $p < .01$, respectively). Mean proportion correct, plotted as a function of block, is presented in Figure 34. There is considerable improvement in all four conditions from Block 1 to Block 3. Also, accuracy in the Whole Task condition was quite high, ranging from 0.91 to 0.97 proportion correct; this was a 0.11 point difference compared to Alternating, the next closest condition. Accuracies in the Alternating and Instruction First conditions were remarkably similar across all blocks of the retention session, while that of the Instruction Last condition remained inferior.

Another question of interest concerned the retention of skill across 65 days without practice. Once again, speed and accuracy were evaluated. A 4 x 2 factorial ANOVA was performed analyzing total time as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and session (transfer and retention). A summary of this analysis is reported in Table 63; mean scores are presented in Table 64. The main effect of training condition was statistically significant ($F(3, 20) = 3.73$, $p < .05$); however, the significant two-way interaction (session x training condition) is of greater interest. Figure 35 portrays this interaction. This figure shows that there is little difference, at transfer, between those conditions in which participants received training on the dispatching task (Alternating and Whole Task). Also, little difference is found also between conditions in which participants received no training on the dispatching task (Instruction First and Last). There is a noticeable difference, however, between those who did and did not receive training on the dispatching task. Furthermore,

Table 62. Summary of Analysis of Variance for Dispatching Task, Retention Session: Proportion Correct

Retention

Source	df	MS	F
condition	3	0.158	3.26*
subjects w/in condition	20	0.048	
block	2	0.038	10.41**
block x condition	6	0.002	
block by subjects w/in condition	40	0.004	

* $p < .05$

** $p < .01$

Table 63. Summary of Analysis of Variance for Dispatching Task, Transfer versus Retention: Total Time to Reach Decision

Source	df	MS	F
condition	3	541034.56	3.73*
subjects w/in condition	20	144879.33	
session	1	60244.80	1.97
session x condition	3	178855.87	5.86**
session by subjects w/in condition	20	30517.06	

* $p < .05$

** $p < .01$

Table 64. Mean Total Time to Reach Decision (seconds) at Transfer and Retention

Session	Instruction First	Instruction Last	Alternating	Whole Task
Transfer	20.39 (15.97)	23.46 (17.11)	11.62 (6.41)	14.18 (9.51)
Retention	20.15 (25.68)	19.69 (13.63)	13.99 (8.26)	20.99 (18.74)

NOTE: Means are on top; standard deviations are in parentheses underneath.

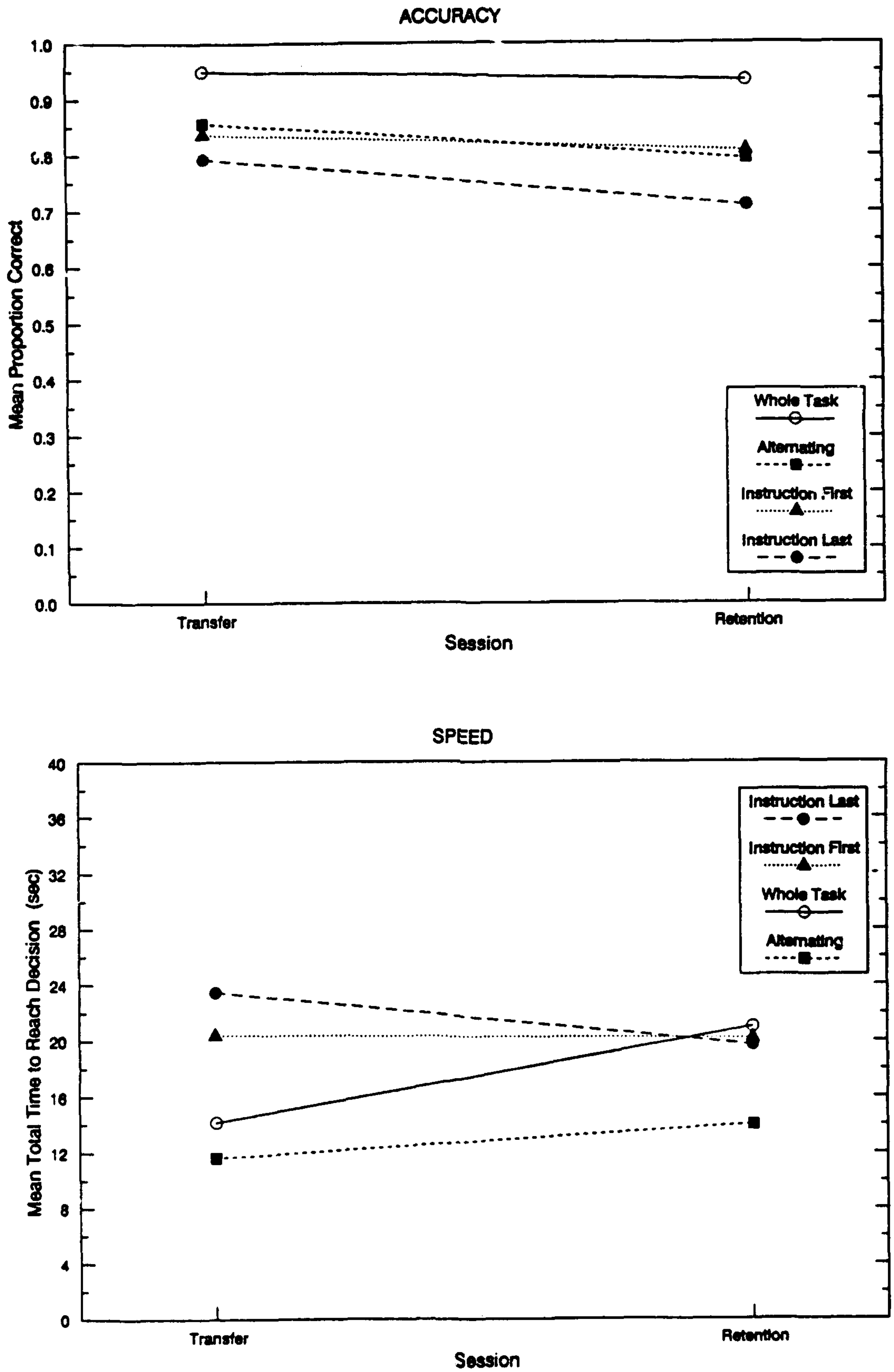


Figure 35. Speed and Accuracy as Functions of Session

performance is stable or even improves slightly across the retention interval for those without dispatching task training. By contrast, a modest increase in the Alternating condition and a marked increase in the Whole Task condition in total time performance were observed for the other two conditions. In fact, the differences between the three worst conditions were neglible and performance in the Whole Task condition was inferior to all others.

The results are somewhat different for the accuracy data. Accuracy performance (proportion correct) as a function of training condition (Instruction First, Instruction Last, Alternating, and Whole Task) and session (transfer and retention) was analyzed in a 4 x 2 factorial ANOVA. A summary of this analysis is presented in Table 65; the means by condition are presented in Table 66. Both the main effects of training condition and session were significant. Mean proportion correct, plotted as a function of session for each of the training conditions, is presented in Figure 35. Quite modest declines in accuracy were observed in the Instruction First and Whole Task conditions--from 0.84 to 0.81 proportion correct and from 0.95 to 0.94, respectively. However, the decline in accuracy was more striking in the Instruction Last and Alternating conditions, yielding a diminution from 0.79 to 0.71 proportion correct and from 0.86 to 0.79 for the two training conditions, respectively. While there was no noticeable difference between the Alternating and Instruction First conditions at either transfer or retention, performance in the Whole Task condition was noticeably superior. Performance in the Instruction Last condition was noticeably inferior in both sessions.

Subjective Workload. Mean workload ratings for each training condition are presented in Table 67 as a function of workload dimension and training or retention session.

Table 65. Summary of Analysis of Variance for Dispatching Task, Transfer versus Retention: Proportion Correct

Source	df	MS	F
condition	3	0.233	4.59**
subjects w/in condition	20	0.051	
session	1	0.086	5.51*
session x condition	3	0.008	
session by subjects w/in condition	20	0.016	

* $p < .05$

** $p < .01$

Table 66. Mean Proportion Correct at Transfer and Retention

Session	Instruction First	Instruction Last	Alternating	Whole Task
Transfer	0.84 (0.10)	0.79 (0.09)	0.86 (0.10)	0.95 (0.06)
Retention	0.81 (0.19)	0.71 (0.16)	0.79 (0.09)	0.94 (0.07)

NOTE: Means are on top; standard deviations are in parentheses underneath.

Table 67. Average Workload Ratings for Each Training Condition

Training Condition	Dimension	Day 1	2	3	4	5	65*
Instruction First	Mental Demand	73	80	68	63	57	82
	Physical Demand	22	29	40	34	29	30
	Temporal Demand	63	77	66	64	52	66
	Effort	55	71	71	69	57	70
	Performance	49	41	29	43	48	48
	Frustration	49	56	46	64	48	58
	Overall	60	68	58	64	54	67
Instruction Last	Mental Demand	80	87	69	61	85	74
	Physical Demand	16	14	13	13	13	13
	Temporal Demand	58	78	57	32	47	63
	Effort	76	81	69	63	80	71
	Performance	32	45	31	45	47	53
	Frustration	55	63	46	37	61	56
	Overall	66	76	60	53	69	67
Alternating	Mental Demand	60	77	78	79	59	68
	Physical Demand	18	35	31	38	27	30
	Temporal Demand	58	76	68	71	70	62
	Effort	58	74	74	73	67	70
	Performance	46	48	51	48	55	46
	Frustration	60	60	59	69	51	49
	Overall	59	70	71	73	64	65
Whole Task First	Mental Demand	72	83	52	55	48	83
	Physical Demand	13	43	20	18	18	22
	Temporal Demand	51	57	53	47	31	51
	Effort	61	80	58	58	40	72
	Performance	58	46	24	24	30	41
	Frustration	49	51	32	34	27	48
	Overall	62	68	45	47	39	64

*Day 65 ratings were collected when participants returned for the retention phase of the experiment.

Mean workload ratings from the transfer session (day 5) and the retention session (Day 65) were subjected to a factorial ANOVA. A separate Training Condition x Session (4 x 2) factorial ANOVA was used for each workload dimension and overall workload ratings.

A main effect of session (a significant change in subjective workload at the retention phase) was found for the mental demand ($F(1,20) = 7.10$, $p < .05$), temporal demand ($F(1,20) = 4.18$, $p < .06$), Effort ($F(1,20) = 4.91$, $p < .05$), and overall workload dimensions ($F(1,20) = 4.91$, $p < .05$). On the average, participants estimated workload higher during the retention session. A Training Condition X Session interaction (indicative of differential changes in perceived workload) was found for the Mental Demand ($F(3,20) = 3.47$, $p < .05$) and Effort dimensions ($F(3,20) = 3.95$, $p < .05$).

The pattern of change in perceived workload from transfer to retention is complex. While, on average, workload increased at retention, it actually decreased in some training conditions on particular dimensions. For example, participants in the Instruction Last condition perceived mental demand and effort to be high at transfer (with ratings of 85 and 80, respectively). At retention, there was an overall increase in perceived mental demand and effort for participants in the Instruction First, Alternating, and Whole Task conditions, while perceived workload for those in the Instruction Last condition declined slightly. This effect was inconsistent across workload dimensions. At retention the Alternating condition showed a decline in perceived workload for temporal demand, while the other three conditions increased in perceived workload for this dimension.

It is posited that participants in the Instruction Last condition, having learned the declarative information

without knowledge of the dispatching task context, perceived the mental demand and effort required by the task as high. At retention, these participants had learned more about the task, and viewed it as somewhat less demanding. Those in the Alternating condition already had one session of the dispatching task. At transfer, participants were encouraged to improve performance over their previous level. Although accuracy was stressed throughout the experiment, participants might have perceived more temporal demand at transfer (Day 5). However, at retention, participants were encouraged to emphasize accuracy and temporal demand subsequently declined in the Alternating condition.

Workload ratings for each of the 24 participants were correlated with performance on the dispatching task at transfer (Day 5) and at retention (Day 65). These correlations are displayed in Table 68. The correlations between accuracy, total time to reach a decision, and each of the workload dimensions are presented in the first two columns of Table 68. The workload dimensions of mental demand, temporal demand, and frustration were negatively correlated with accuracy on the last day of training ($p < .05$), as was the overall workload measure. Participants who were performing better at the dispatching task also viewed that task as less frustrating and less demanding. Participants who had more difficulty with the task viewed it as more demanding and frustrating. No significant correlations were obtained between workload and the total time required to reach decisions.

Correlations between the workload dimensions and performance at retention are also presented in Table 69. At retention, only the frustration and performance dimensions were negatively correlated with accuracy ($p < .05$). While not statistically significant, reasonably high positive correlations ($r = .38$ $p < .06$) were obtained between mental

Table 68. Correlation Matrix of Workload Measures, Accuracy, and Total Time to Reach Decision, in the Transfer Session (Day 5)

Variable	1	2	3	4	5	6	7	8	9
1. Accuracy	.								
2. Total Time	-.31	.							
3. Mental Demand	-.53	.34	.						
4. Physical Demand	-.25	-.03	.27	.					
5. Temporal Demand	-.26	-.29	.46	.57	.				
6. Effort	-.69	.27	.85	.33	.58	.			
7. Performance	-.37	-.11	.32	.46	.68	.44	.		
8. Frustration	-.64	.27	.65	.50	.49	.67	.61	.	
9. Overall Workload	-.61	.12	.87	.44	.73	.89	.66	.82	.

Note: With an N of 24 participants, a correlation greater than 0.39 is significant at the .05 level. Significant correlations are displayed in bold typeface.

Table 69. Correlation Matrix of Workload Measures, Accuracy, and Total Time to Reach Decision in the Retention Session (Day 65)

Variable	1	2	3	4	5	6	7	8	9
1. Accuracy	.								
2. Total Time	.24	.							
3. Mental Demand	.06	.38	.						
4. Physical Demand	.15	-.11	-.07	.					
5. Temporal Demand	-.01	.05	.25	.30	.				
6. Effort	-.13	.28	.69	.02	.25	.			
7. Performance	-.52	-.23	-.03	.14	.06	.03	.		
8. Frustration	-.53	.38	.06	.08	.08	.30	.44	.	
9. Overall Workload	-.28	.18	.71	.13	.50	.83	.39	.45	.

Note: With an *N* of 24 participants, a correlation greater than 0.39 is significant at the 0.05 level. Significant correlations are displayed in bold typeface.

demand, frustration, and the total time required to reach decisions. The change in the pattern of significant workload/performance correlations at retention is due to the general increase in perceived workload. At retention, all four training conditions viewed the dispatching task as mentally and temporally demanding, regardless of their performance.

Discussion

Training Data. In the Training Phase of this investigation, high-performance skill development as a function of whole-task versus part-task training was examined. Part-task training was employed to facilitate the development of critical, declarative knowledge components requisite to perform the whole task.

The performance of individuals who received contextually relevant instructions regarding operator and destination names associated with their respective class acronyms was markedly superior to the performance of individuals who were instructed only that the names would be used later in a more complex task. This effect is interpretable in terms of the association of input with existing nodes in the semantic memory network. The acquisition of declarative knowledge usually does not involve something entirely new; rather it involves adding more details to a well-developed conceptual network (Glass and Holyoak, 1986). In other words, memory is partly a by-product of understanding: people do not understand a description fully unless they can imagine a concrete example of what is being described (Branford and Johnson, 1972; McFarland, 1986).

The present experimental paradigm was designed to minimize the role of any previously existing conceptual knowledge (i.e., prior associations with the experimental

stimuli). This paradigm allowed us to examine the development of a high-performance skill which places heavy demands on the working memory and memory scanning components of the human information-processing system. The provision of contextually relevant information may have facilitated study and RT performance in the Instruction First and Alternating conditions by allowing the assimilation of to-be-remembered input into an existing conceptual structure. In this manner, the association of operator and destination names with their respective class acronyms was facilitated.

It is important to note that, by the end of training, performance in all part-task conditions was characterized by a high level of accuracy. Clearly, all participants in these conditions effectively acquired declarative knowledge components integral to performance on the dispatching task. The part-task training data demonstrate the need to provide instructions regarding the ultimate application of to-be-learned material prior to providing part-task training.

Acquisition of dispatching task skill in the Whole Task condition was characterized by dramatic improvements in total time performance early in training, with more modest improvement exhibited later in training. This pattern of data is suggestive of power functions found typically in the skill-acquisition domain (Newell and Rosenbloom, 1981). These individuals exhibited steady improvement in accuracy until reaching ceiling toward the end of training. The performance level obtained in the Whole Task condition served as a reasonable index with which performance in the part-task conditions could be compared. Furthermore, performance in the Whole Task condition was characterized by an increased speed of decisions, and a reduction in both the use of help screens and the number of keystrokes required to complete the task (Appendix C).

Training on the dispatching task in the Alternating condition, in which participants alternately practiced the memory-search and dispatching tasks, was characterized by a dramatic improvement in total time and reasonable improvement in accuracy. Accuracy performance rivaled that seen early in Whole Task training. In contrast, total time performance in the Alternating condition was remarkably faster than that exhibited early in Whole Task training, and just as fast as that demonstrated by Whole Task participants in the middle of training. It is clear that part-task training greatly facilitated development of the dispatching task skill, producing dramatic savings in total time performance.

Transfer Data. At transfer, dispatching-task-trained participants (i.e., participants in the Alternating and Whole Task conditions) performed better than part-task-only-trained participants (i.e., Instruction First and Last conditions). The effectiveness of whole-task training, however, was relatively small compared to the effectiveness of providing part-task training with contextually relevant instructions. Dispatching-task performance of participants who received contextually relevant instructions regarding the application of the to-be-learned material was markedly superior at transfer to performance of participants in the Instruction Last condition who were instructed only that the to-be-learned material would be used later. The advantages of whole-task-relevant instruction persisted throughout the transfer session. Data from the NASA-TLX subjective workload measure were generally consistent with the performance data.

Retention Data. A comparison of transfer and retention performance revealed considerable variability across the different training conditions. Whole Task participants were able to maintain a consistently high degree of accuracy

across the retention interval; however, the cost of this maintenance was a reduction in total time performance which declined noticeably. In the Alternating condition, there were modest declines in both accuracy and total time performance. There was no noticeable decline in speed for the memory-search only conditions (i.e., Instruction First and Last). The pattern was different for the retention of accuracy; however, the absence of contextually relevant instructions resulted in a noticeable reduction of accuracy in the Instruction Last condition, while in the Instruction First condition accuracy declined only slightly.

In comparing performance among the various training conditions at retention, the relative rankings of the conditions were generally maintained across the retention interval. First, performance in the Whole Task condition was strikingly superior in terms of accuracy. Conversely, performance in the Alternating condition was markedly superior in terms of speed. Again, the lack of contextually relevant instruction on the dispatching task resulted in the poorest accuracy performance for the Instruction Last condition. All conditions improved across blocks in the retention phase, both in accuracy and total time.

In this investigation we examined the effects of part- and whole-task procedures on skill acquisition and retention in a relatively complex "strategic planning" task. Most significantly, the part-task training data clearly illustrate the necessity of providing instructions regarding the ultimate application of to-be-trained material prior to providing training. These data dramatically reveal the value of simplified part-task training for facilitating the development of declarative knowledge underlying the effective performance of complex decision-making tasks. Although Whole Task performance was generally

superior--given that maintaining a high level of accuracy is ultimately the most important criterion of performance--the marked superiority of the Alternating condition in total time performance is suggestive. Future investigations may help to elucidate a methodology by which improvement in accuracy, without a concommitant reduction in speed, might be obtained.

VII. LESSONS LEARNED AND AUGMENTED PROCESSING PRINCIPLES

Toward a Formal Theory of Part-Task Training

This phase of the research effort has added substantially to our understanding of skill development, especially for tasks that rely on visual-search components for successful performance. First, in concurrence with other researchers in the field (e.g., Fisher, 1986; Fisher and Tanner, in press; Fisk and Rogers, 1991; Shiffrin, 1991), we have documented that performance improvement in visual search is the result of multiple (perhaps interacting) learning mechanisms. The present data, and data from past investigations, suggest the direction that a formal theory of part-task training should take.

Our research, and research from numerous other laboratories, indicate that the marked performance improvements that comes with practice when stimuli are consistently mapped to responses are primarily (though not solely) the consequence of a change in the mode of information processing. In memory search, the data (Fisk and Rogers, 1991; Schneider and Detweiler, 1988; Shiffrin and Schneider, 1977) suggest that individuals shift from a slow serial search (controlled processing) to a fast parallel search (automatic processing).

In visual search, Fisher (1982, 1984, 1986; Fisher and Tanner, in press) has recently suggested that individuals switch from a slow, random processing of target features at the beginning of practice to a fast, optimal (though attentive) processing of target features at the end of practice. Several investigators have argued that this alternative to automatism can explain several results from the visual search literature which are otherwise difficult to interpret (e.g., Czerwinski, 1988; Shiffrin, 1991).

Finally, Lawless and Eggemeier (1990) have shown that the hypothesis that subjects switch strategies is consistent with performance improvements in complex, operational tasks as well as simple, laboratory tasks.

However, although a switch in strategies seems necessary, it does not seem sufficient to account for high-performance skill in visual search. Fisk and Rogers (1991; Rogers, 1991; Rogers and Fisk, in press; Section II and III) clearly demonstrate that extended consistent training in visual search can lead to development of efficient optimal search strategies as well as the training of attention (development of an automatic response). The data presented in this report (Sections IV and V) also suggest that the training environment can and will have a major effect on either the development or use of optimal search strategies. In addition, Lee, Rogers, and Fisk (1991) have shown that "simplification" techniques which increase the distribution of practice on specific target-distractor pairings can greatly affect optimal search development.

We can conclude that processing requiring memory search and processing requiring visual search will be influenced by different learning mechanisms. In memory search, consistent training will lead to associative learning of the memory elements and produce an "automatic category" response. Changes in strategic processing may also occur; however, although these strategic changes increase efficiency, strategic processing still requires attention. In visual search, performance improvement is due to learning general search strategies; learning optimal, strategic search patterns; and developing target-distractor strength differentiation. In visual search, especially when speed of response is the dependent measure, performance may asymptote but learning will continue.

Individuals developing part-task training must be cognizant of the above difference in task components and potential learning mechanisms. Although training designers must attend to "what is consistent" about a task, further attention must be given to task factors in order to optimize training. Hence, a formal theory of part-task training must be able to predict optimization of shifts in strategic processing and identify procedures to maximize automatism for both memory and visual-search dependent processing. Utilizing such an approach, training programs can be designed to develop optimal initial training and optimal higher stages of training. Such a formal approach to part-task training should also make it possible to predict when it is optimal to promote a given trainee to the next higher stage of training.

Augmented Processing Principles

An important outcome of this research program is the continuing opportunity to specify "processing principles." Processing principles illustrate human performance guidelines that have been important for the development of "knowledge engineering" required for understanding and developing training programs for complex operational tasks. The guidelines should be used when the to-be-trained situation is well understood. The previous principles of human performance have been described in Fisk et al. (1987) and more recently in Fisk, Rogers et al. (1991).

Based on the present work, we are again in a position to add to these human performance guidelines. The additional guidelines allow further specification of human performance principles for determining performance limits and training program design for high-performance-skills training in complex tasks.

Phase 3 Processing Principles

1. In pure visual search, an individual's performance improvement is guided by the same factors (learning general search strategies and then optimal, stimulus-specific search strategies) early in practice for both CM and VM training conditions. Qualitative differences between CM and VM performance are seen late in practice and generally when the implementation of the attentive optimal search is difficult. (Czerwinski, 1988; Rogers, 1991; Section II)

2. Improvement on pure memory search tasks (or task components) is extremely fast and asymptotic performance is reached much faster than for pure visual search tasks. (Section III)

3. Performance improvement is driven by different mechanisms in memory and visual search. However, the learning in visual search transfers to memory search. The learning in memory search does not transfer well to visual search. (Section III)

4. The need to inhibit one automatic process does not imply that other automatic processes performed within the same context will be disrupted. (Section IV)

VIII. REFERENCES

Ackerman, P. L. (1984). A theoretical and empirical investigation of individual differences in learning: A synthesis of cognitive ability and information processing perspectives. Unpublished doctoral dissertation, University of Illinois, Urbana.

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. Intelligence, 10, 101-139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Ackerman, P. L. (1989). Individual differences and skill acquisition. In P. L. Ackerman, R. J. Sternberg, and R. Glaser (Eds.), Learning and individual differences (pp. 165-217). New York: W. H. Freeman.

Ackerman, P. L., and Schneider, W. (1985). Individual differences in automatic and controlled information processing. In R. F. Dillon (Ed.), Individual differences in cognition (Vol. 2, pp. 35-66). Orlando, FL: Academic Press.

Adams, J. A. (1957). The relationship between certain measures of ability and acquisition of a psychomotor response. Journal of General Psychology, 56, 121-134.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41-74.

Alvares, K. M., and Hulin, C. L. (1972). Two explanations of temporal changes in ability-skill relationships: A literature review and a theoretical analysis. Human Factors, 14, 293-308.

Alwin, D. W. (1988). Structural equation models in research on human development and aging. In K. W. Schaie, R. T. Campbell, W. Meredith, and S. C. Rawlings (Eds.), Methodological issues in aging research (pp. 71-170). New York: Springer.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Atkinson, R. C., Holmgren, J. E., and Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. Perception and Psychophysics, 6, 321-326.

Baddeley, A. D. (1986). Working Memory. Oxford Psychological Series No. 11. Oxford: Clarendon Press.

Bader, W., and Burt, D. S. (1983). Miller Analogies Test. New York: Arco Publishing.

Battig, W. F., and Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 80, (Whole).

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley and Sons.

Botwinick, J., and Storandt, M. (1974). Memory related functions and age. Springfield, IL: Charles C. Thomas.

Bransford, J. D., and Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-721.

Bryan, W. L., and Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. Psychological Review, 6, 345-375.

Carlson, R. A., Sullivan, M. A., and Schneider, W. (1989). Practice and working memory effects in building procedural skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 517-526.

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon and R. R. Schmeck (Eds.), Individual differences in cognition (Vol. 1, pp. 1-33). New York: Academic Press.

Carroll, J. B. (1988). Individual differences in cognitive functioning. In R. C. Atkinson, R. J. Herrnstein, and R. D. Luce (Eds.), Stevens' handbook of experimental psychology (pp. 813-862). New York: Wiley and Sons.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.

Cheng, P.W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. Psychological Review, 92, 414-423.

Collen, A., Wickens, D. D., and Daniele, L. (1975). The interrelationship of taxonomic categories. Journal of Experimental Psychology: Human Learning and Memory, 1, 629-633.

Cook, T. D., and Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin Co.

Corballis, M. C. (1965). Practice and the simplex. Psychological Review, 72, 399-406.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix and H. Hagendorf (Eds.), Human memory and cognitive abilities (pp. 409-422). Amsterdam: Elsevier.

Czerwinski, M. P. (1988). Differences between visual and memory search: Implications for models of attention. Unpublished doctoral dissertation, Indiana University.

Dodd, D. H., and Schultz, R. F. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance designs. Psychological Bulletin, 79, 391-395.

Dumais, S. T. (1979). Perceptual learning in automatic detection: Processes and mechanisms. Unpublished doctoral dissertation, Indiana University.

Duncan, J. (1986). Consistent and varied training in the theory of automatic and controlled processing. Cognition, 23, 279-284.

Duncan, J., and Humphreys, G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96, 433-458.

Durso, F. T., Cooke, N. M., Breen, T. J., and Schvaneveldt, R. W. (1987). Is consistent mapping necessary for high-speed search? Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 223-229.

Dwyer, P. S. (1937). The determination of the factor loading of a given test from the known factor loadings of other tests. Psychometrika, 2, 173-178.

Eggemeier, F. T., Fisk, A. D., Robbins, R. J., and Lawless, M. T. (1988). Application of automatic/controlled processing theory to training tactical command and control skills: II. Evaluation of a task analytic methodology. Proceedings of the Human Factors Society 32nd Annual Meeting. Santa Monica, CA: Human Factors Society.

Ekstrom, R. B., French, J. W., and Harman, H. (1979). Cognitive factors: Their identification and replication. Multivariate Behavior Research Monographs, 2, 1-84.

Ekstrom, R. B., French, J. W., Harman, H., and Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Everitt, B. S. (1984). An introduction to latent variable models. London: Chapman and Hall.

Ferguson, G. A. (1954). On learning and human ability. Canadian Journal of Psychology, 8, 95-112.

Ferguson, G. A. (1956). On transfer and the abilities of man. Canadian Journal of Psychology, 10, 121-131.

Fisher, D. L. (1982). Limited channel models of automatic detection: Capacity and scanning in visual search. Psychological Review, 89, 662-692.

Fisher, D. L. (1984). Central capacity limits in consistent mapping visual search tasks: Four channels or more? Cognitive Psychology, 16, 449-484.

Fisher, D. L. (1986). Hierarchical models of visual search: Serial and parallel processing. Paper presented at the annual meeting of the Society for Mathematical Psychology, Cambridge, MA.

Fisher, D. L., Duffy, S. A., Young, C., and Pollatsek, A. (1988). Understanding the central processing limit in consistent-mapping visual search tasks. Journal of Experimental Psychology: Human Perception and Performance, 14, 253-266.

Fisher, D. L., and Fisk, A. D. (1991). A formal theory of part-task training: Individualizing the learning and retention of high demand skills. Unpublished manuscript.

Fisher, D. L., and Tanner, N. S. (in press). Optimal symbol selection: A semi-automated procedure. Human Factors.

Fisk, A. D., Ackerman, P. L., and Schneider, W. (1987). Automatic and controlled processing theory and its application to human factors problems. In P. A. Hancock (Ed.), Human factors psychology (pp. 159-197), New York: North-Holland Publishing Co.

Fisk, A. D., and Eggemeier, F. T. (1988). Application of automatic/controlled processing theory to training tactical command and control skills: 1. Background and task analytic methodology. Proceedings of the Human Factors Society 32nd Annual Meeting. Santa Monica, CA: Human Factors Society.

Fisk, A. D., Giambra, L. M., Rogers, W. A., and Rosenberg, E. H. (1990, March). Efficiency of associative and priority learning in young, young-old, and old adults. Presented at the third Cognitive Aging Conference, Atlanta.

Fisk, A. D., Hodge, K. A., Lee, M. D., and Rogers, W. A. (1990). Automatic information processing and high-performance skills: Acquisition, transfer, and retention (AFHRL-TR-89-69, AD-A221 744). Wright-Patterson AFB, OH: Logistics and Human Factors Division, Air Force Human Resources Laboratory.

Fisk, A. D., Lee, M. D., and Rogers, W. A. (in press). Recombination of automatic processing components: The effects of transfer, reversal, and conflict situations. Human Factors.

Fisk, A. D., and Rogers, W. A. (1988). The role of situational context in the development of high-performance skills. Human Factors, 30, 703-712.

Fisk, A. D., and Rogers, W. A. (1991). Toward an understanding of age-related memory and visual search effects. Journal of Experimental Psychology: General, 120, 131-149.

Fisk, A. D., Rogers, W. A., and Lee, M. D. (1991). Learning and performance improvements as a function of amount and type of practice: Issues of massed vs. distributed practice and task-relevant vs. stimulus-specific learning. Unpublished manuscript.

Fisk, A. D., Rogers, W. A., Lee, M. D., Hodge, K. A., and Whaley, C. J. (1991). Automatic information processing and high-performance skills: Principles of consistency, part-task training, context, retention, and complex task performance (AFHRL-TR-90-84). Wright-Patterson AFB, OH: Logistics and Human Factors Division, Air Force Human Resources Laboratory.

Fisk, A. D., and Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. Human Factors, 23, 737-750.

Fisk, A. D., and Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 177-195.

Fitts, P. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning (pp. 243-285). New York: Academic Press.

Fitts, P., and Posner, M. I. (1967). Human performance. Belmont, CA: Brooks/Cole.

Flach, J. M. (1986). Within-set discriminations in a consistent mapping search task. Perception and Psychophysics, 39, 397-408.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. American Psychologist, 27, 1017-1032.

Fleishman, E. A. (1975). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

Fleishman, E. A., and Hempel, W. E. (1954). Changes in factor structure of a complex motor test as a function of practice. Psychometrika, 18, 239-252.

Fleishman, E. A., and Hempel, W. E., Jr. (1955). The relation between abilities and improvement with practice in a visual discrimination reaction task. Journal of Experimental Psychology, 49, 301-312.

Giambra, L. M., and Quilter, R. E. (1988). Sustained attention in adulthood: A unique, large-sample, longitudinal, and multi-cohort analysis using the Mackworth clock test. Psychology and Aging, 3, 75-83.

Glass, A. L., and Holyoak, K. J. (1986). Cognition (2nd ed.). New York: Random House.

Guilford, J. P. (1967). _The nature of human intelligence_. New York: McGraw-Hill.

Guilford, J. P. (1985). The structure of intellect model. In B. B. Wolman (Ed.), _Handbook of intelligence_ (pp. 225-266). New York: Wiley and Sons.

Gustaffson, J. (1984). A unifying model for the structure of intellectual abilities. _Intelligence_, _8_, 179-203.

Halff, H. M., Hollan, J. D., and Hutchins, E. L. (1986). Cognitive science and military training. _American Psychologist_, _41_, 1131-1139.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), _Human Mental Workload_ (pp. 139-183). Amsterdam: North Holland.

Henry, R. A., and Hulin, C. L. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. _Journal of Applied Psychology_, _72_, 457-462.

Hertzog, C. (1985). Applications of confirmatory factor analysis to the study of intelligence. In D. K. Detterman (Ed.), _Current topics in human intelligence: Research methodology_ (Vol. 1, pp. 59-97). Norwood, NJ: Ablex.

Hertzog, C. (1990). On the utility of structural equation models for developmental research. In P. B. Baltes, D. Featherman, and R. M. Lerner (Eds.), _Life-span development and behavior_ (Vol. 10, pp. 257-290). Hillsdale, NJ: Erlbaum.

Hertzog, C., Raskind, C. L., and Cannon, C. J. (1986). Age-related slowing in semantic information processing speed: An individual differences analysis. _Journal of Gerontology_, _41_, 500-502.

Hertzog, C., and Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. _Psychology and Aging_, _1_, 159-171.

Hill, P. W. (1987). Modeling the hierarchical structure of learning. In P. Cuttance and R. Ecob (Eds.), _Structural modeling by example: Applications in educational, sociological, and behavioral research_ (pp. 65-85). Cambridge: Cambridge University Press.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. Perception and Psychophysics, 23, 1-11.

Hoffman, J. E. (1979). A two-stage model of visual search. Perception and Psychophysics, 25, 319-327.

Horn, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. I. M. Craik and S. Trehub (Eds.), Aging and cognitive processes (pp. 237-278). New York: Plenum Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), Handbook of intelligence (pp. 267-300). New York: Wiley and Sons.

Horn, J. L., and Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. Acta Psychologica, 26, 107-129.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. Journal of Gerontology, 35, 225-231.

Humphreys, L. G. (1960). Investigation of the simplex. Psychometrika, 25, 313-323.

Humphreys, L. G. (1979). The construct of general intelligence. Intelligence, 3, 105-120.

Hunt, E., Frost, N., and Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), Psychology of learning and motivation (Vol. 7). New York: Academic Press.

Hunt, E., Lunneborg, C., and Lewis, J. (1975). What does it mean to be high verbal? Cognitive Psychology, 7, 194-227.

James, L. R., Mulaik, S. A., Brett, J. M. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills, CA: Sage Press.

Joreskog, K. G. (1970). Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, 23, 121-145.

Joreskog, K. G., and Sorbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D. J. Aigner and A. S. Goldberger (Eds.), Latent variables in socio-economic models (pp. 631-639). Amsterdam: North-Holland.

Joreskog, K. G., and Sorbom, D. (1986). LISREL VI - Analysis of linear structural relationships by the method of maximum likelihood (4th ed.). Mooresville, IN: Scientific Software, Inc.

Kenny, D. A. (1979). Correlation and causality. New York: John Wiley and Sons.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kessler, R. C., and Greenberg, D. F. (1981). Linear panel analysis: Models of quantitative change. New York: Academic Press.

Kirk, R. E. (1982). Introduction to basic concepts in experimental design (Chapter 1). In Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Kramer, A., Schneider, W., Fisk, A. D., and Donchin, E. (1986). The effects of practice and task structure on components of the event-related brain potential. Psychophysiology, 23, 33-47.

Kristofferson, M. W. (1977). The effects of practice with one positive set in a memory scanning task can be completely transferred to a different positive set. Memory and Cognition, 5, 177-186.

Kyllonen, P. C., Tirre, W. C., and Christal, R. E. (1991). Knowledge and processing speed as determinants of associative learning. Journal of Experimental Psychology: General, 120, 57-79.

Kyllonen, P. C., and Woltz, D. J. (1989). Role of cognitive factors in the acquisition of cognitive skill. In R. Kanfer, P. L. Ackerman, and R. Cudeck (Eds.), Abilities, motivation, and methodology (pp. 239-280). Hillsdale, NJ: Erlbaum.

LaBerge, D., and Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Labouvie-Vief, G., Frohring, W. R., Baltes, P. B., and Goulet, L. R. (1973). Changing relationships between recall performance and abilities as a function of stage of learning and timing of recall. Journal of Educational Psychology, 64, 191-198.

LaPointe, L. B., and Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 1118-1133.

Lawless, M. T., and Eggemeier, F. T. (1990, May) Automatic processing and complex visual search. Paper presented at the Fifth Mid-Central Ergonomics Meeting. Dayton, OH.

Lee, M. D., Rogers, W. A., and Fisk, A. D. (1991). Contextual change and skill acquisition in visual search: Does the rate of change affect performance? Proceedings of the Human Factors Society 35th Annual Meeting. Santa Monica, CA: Human Factors Society.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Logan, G. D., and Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practive necessary to produce automaticity? Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 177-193.

Maccoby, E. E., and Jacklin, C. N. (1974). Intellectual abilities and cognitive styles (Chapter 3, pp. 63-133). In The psychology of sex differences. Stanford: Stanford University Press.

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, 89, 483-506.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of intelligence. Baltimore, MD: Williams & Wilkins.

McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart and J. L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1, pp. 3-44). Cambridge, MA: MIT Press.

McFarland, A. D. (1986). Cognitive effort during reading: Effects of topic knowledge. Unpublished doctoral dissertation, University of New Mexico, Albuquerque.

Minton, H. L., and Schneider, F. W. (1980). Sex differences. In Individual differences (pp. 265-329). Monterey, CA: Brooks/Cole.

Myers, G. L., and Fisk, A. D. (1987). Training consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Neches, R., Langley, P., and Klahr, D. (1987). Learning, development, and production systems. In D. Klahr, P. Langley, and R. Neches (Eds.), Production system models of language and development (pp. 1-54). Cambridge, MA: MIT Press.

Newell, A., and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Peterson, J., and Barlow, M. C. (1928). The effects of practice on individual differences. The yearbook of the National Society for the Study of Education. Chicago: National Society for the Study of Education.

Prinz, W. (1979). Locus of the effect of specific practice in continuous visual search. Perception and Psychophysics, 25, 137-142.

Rabbitt, P. M. A., Cumming, G., and Vyas, S. (1979). Improvement, learning and retention of skill at visual search. Quarterly Journal of Experimental Psychology, 31, 441-459.

Raven, J. C., Court, J. H., and Raven, J. (1977). Raven's progressive matrices and vocabulary scales. New York: Psychological Corporation.

Rogers, W. A. (1989). Target and distractor learning in visual search: Age-related differences. Unpublished master's thesis, Georgia Institute of Technology, Atlanta, GA.

Rogers, W. A. (1991). An analysis of ability/performance relationships as a function of practice and age. Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta, GA.

Rogers, W. A., and Fisk, A. D. (in press). Are age differences in consistent-mapping visual search due to feature learning or attention training? Psychology and Aging.

Rumelhart, D. E., and McClelland, J. L. (1987). Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1). Cambridge, MA: MIT Press.

Ryan, C. (1983). Reassessing the automatic-control distinction: Item recognition as a paradigm case. Psychological Review, 90, 171-178.

Salthouse, T. A. (1985). A theory of cognitive aging. Amsterdam: North-Holland.

Salthouse, T. A., and Babcock, R. L. (1991). Decomposing adult age differences in working memory. Developmental Psychology, 27, 517-527.

Salthouse, T. A., Babcock, R. L., and Shaw, R. L. (1991). Effects of adult age on structural and operational capacities in working memory. Psychology and Aging, 6, 118-127.

Salthouse, T. A., and Mitchell, D. R. D. (1990). Effects of age and naturally occurring experience on spatial visualization performance. Developmental Psychology, 26, 845-854.

Salthouse, T. A., and Somberg, B. L. (1982). Skilled performance: Effects of adult age and experience on elementary processes. Journal of Experimental Psychology: General, 111, 176-207.

Schaie, K. W. (1984). Historical time and cohort effects. In K. A. McCluskey and H. W. Reese (Eds.), Life span developmental psychology: Historical and generational effects (pp. 1-15). Orlando, FL: Academic Press.

Schaie, K. W., and Hertzog, C. (1985). Measurement in the psychology of adulthood and aging. In J. E. Birren and K. W. Schaie (Eds.), Handbook of the psychology of aging (2nd ed., pp. 61-92). New York: Van Nostrand Reinhold.

Schneider, W. (1985). Toward a model of attention and the development of automaticity. In M. I. Posner and O. S. Martin (Eds.), Attention and Performance Volume XI (pp. 475-492). Hillsdale, N.J.: Erlbaum.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. Behavior Research Methods, Instrumentation, and Computers, 20, 206-217.

Schneider, W., and Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 21, pp. 53-118). New York: Academic Press.

Schneider, W., and Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload

modeling in a connectionist/control architecture. Human Factors, 30, 539-566.

Schneider, W., and Fisk, A. D. (1982a). Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 31, 160-168.

Schneider, W., and Fisk, A. D. (1982b). Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 261-278.

Schneider, W., and Fisk, A. D. (1983). Attentional theory and mechanisms for skilled performance. In R. Magil (Ed.), Memory and control in motor behavior. New York: North-Holland Publishing Co.

Schneider, W., and Fisk, A. D. (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 1-15.

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, attention. Psychological Review, 84, 1-66.

Schneider, W., and Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors. Psychological Review, 92, 424-428.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, and R. D. Luce (Eds.), Stevens' handbook of experimental psychology (pp. 739-811). New York: Wiley and Sons.

Shiffrin, R. M. (1991). Automatization and attention learning in visual search. Presented at the Conference on "Human Error," Indiana University, Bloomington, IN.

Shiffrin, R. M., and Czerwinski, M. P. (1988). A model of automatic attention attraction when mapping is partially consistent. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 562-569.

Shiffrin, R. M., and Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 111-140). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Shiffrin, R. M., and Schneider, W. (1984). Automatic and controlled processing revisited. Psychological Review, 91, 269-276.

Siegel, S. (1956). Nonparametric statistics: For the behavioral sciences. New York: McGraw-Hill.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Sternberg, R. J. (1985). General intellectual ability. In R. J. Sternberg (Ed.), Human abilities: An information-processing perspective (pp. 5-30). San Francisco: Freeman.

Strayer, D. L., and Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 291-304.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. Journal of Experimental Psychology: Human Perception and Performance, 8, 194-214.

Treisman, A. M., and Gelade, G. (1980). A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Treisman, A., and Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. Psychological Review, 95, 15-48.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. American Psychologist, 30, 128-134.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale: Revised. New York: Psychological Corporation.

Wightman, D. C., and Lintern, G. (1985). Part-task training for tracking and manual control. Human Factors, 27, 267-284.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. Journal of Experimental Psychology: General, 117, 319-331.

Woodrow, H. (1946). The ability to learn. Psychological Review, 53, 147-158.

APPENDIX A. INSTRUCTIONS FOR PERFORMING THE DISPATCHING TASK

Task Instructions

In this task you will perform the duties of a dispatcher. Your task is to assign operators to vehicles which deliver cargo to different destinations. On each trial, you will receive an order containing the following information: 1) the type of cargo to be delivered, 2) the weight of the cargo in kilograms (kg), 3) the vehicle which is available to transport the cargo, and 4) the destination to which the cargo is to be delivered. The names of four vehicle operators are presented on each trial. You must select one operator (the optimal choice out of all four operators) to deliver the cargo.

A set of rules governs the decision-making process for selection of the optimal operator. In order to correctly choose the best operator, you will need to know how the destinations, cargos, vehicles, operators, etc. are classified.

Let's explore the structure of the task in greater detail. First, we'll examine the classification scheme. There are six sets of classes (or categories):

1. CARGO - the type of material to be transported.
There are three classes of cargo: **general purpose** (GP), **liquid** (LQ), and **hazardous** (HZ).

2. WEIGHT - the weight of the cargo to be transported.
There are three classes of cargo weight: **light** (L), **medium** (M), and **heavy** (H).

3. DISTANCE - the distance the cargo must be transported.
There are three classes of distance to destination (**short range** (SR), **medium range** (MR), and **long range** (LR).

4. VEHICLES - the type of vehicle to be used.
There are nine types of vehicles. Vehicles are divided into three principal classes based on the kind of cargo they carry (general purpose, liquid, and hazardous). Each principal class is divided further into three sub-classes based upon weight rating (**light duty**, **medium duty**, and **heavy duty**).

5. DESTINATIONS - the company to which the cargo is to be delivered. Destinations are divided into three principal classes based upon the type of cargo which they receive (general purpose, liquid, or hazardous). Each of these classes is divided further into three sub-classes based upon distance from the shipping terminal (short, medium, or long).

6. OPERATOR LICENSES - the skill level of the driver required to complete a delivery (determined by the type of cargo, and weight of cargo to be transported).
There are nine levels of operator licenses. Licenses are divided into three principal classes based upon the distance the operator is permitted to transport cargo (short, medium, or long range) and the type of cargo to be delivered (general purpose, liquid, and hazardous). Also, each principal class is subdivided into three sub-classes based upon the weight rating of the vehicle the operator is permitted to operate (light duty, medium duty, or heavy duty). The license classification system is a progressive one: an operator with a given license classification is permitted to do all that an operator with a lower license classification can do (more about this later).

RULES

The following rules govern the assignment of operators to deliveries.

DISPATCHING DECISION RULE

Because a drivers' salary is determined by his license class, the operator with the lowest license classification who is qualified to operate the available vehicle is to be given the assignment. Drivers with higher licenses must be paid more. For example, Barney has a license classification of 2.1 (a qualified license) and Olivia has a license classification of 3.2 (a higher class, qualified license). If they are both qualified to do the job then Barney should be given the assignment. This is the rule that operates to minimize cost (i.e., send the operator who is paid the least).

VEHICLE RULES

1. Any vehicle can travel any distance to deliver its cargo. There is no restriction of range for vehicles.

2. A vehicle classified as "light duty" (LD), can carry from 0-1,500 kilograms (kg).
3. A vehicle classified as "medium duty" (MD), can carry from 0-10,000 kg.
4. A vehicle classified as "heavy duty" (HD), can carry any size load, and there is no maximum limitation.
5. A vehicle classified as "general purpose" (GP), can carry <u>only</u> cargo that is classified as general purpose.
6. A vehicle classified as "liquid" (LQ), can carry <u>only</u> cargo that is classified as liquid.
7. A vehicle classified as "hazardous" (HZ), can carry <u>only</u> cargo that is classified as hazardous.

Appendix A (continued)

DESTINATION RULES

1. Any destination can receive any amount (i.e., weight) of cargo. There is no restriction on the amount of cargo received by a destination.
2. A destination classified as "**general purpose**" (GP), can receive only cargo classified as general purpose.
3. A destination classified as "**liquid**" (LQ), can receive only cargo classified as liquid.
4. A destination classified as "**hazardous**" (HZ), can receive only cargo classified as hazardous.
5. A destination classified as "**short range**" (SR), requires that a vehicle must travel between 0 and 80 kilometers (km) to deliver its cargo.
6. A destination classified as "**medium range**" (MR), requires that a vehicle must travel between 81 and 320 km to deliver its cargo.
7. A destination classified as "**long range**" (LR), requires that a vehicle must travel more than 320 km to deliver its cargo.

LICENSE RULES

General Purpose and Short Range

If an operator is classified 1.1, then he or she can:

1) operate vehicles classified as "**general purpose**" and "**light duty**" (GP-LD)

2) only deliver cargo to destinations classified as "**short range**" (SR).

If license = 1.1, then vehicle = GP-LD and destination = SR.

Appendix A (continued,

If an operator is classified 1.2, then he or she can:
1) operate vehicles classified as "medium duty" (MD) in addition to vehicles that a driver with a 1.1 license can operate.

If license = 1.2, then vehicle = GP-LD or GP-MD and destination = SR.

If an operator is classified 1.3, then he or she can:
1) operate vehicles classified as "heavy duty" (HD) in addition to vehicles that a driver with a 1.1 or a 1.2 license can do.

If license = 1.3, then vehicle = GP-LD or GP-MD or GF-HD and destination = SR.

Liquid and Medium Range
If an operator is classified 2.1, then he or she can:
1) operate vehicles classified as "general purpose" and either "light duty" (GP-LD), "medium duty" (GP-MD), or "heavy duty" (GP-HD) plus vehicles classified as "liquid" and "light duty" (LQ-LD)
2) only deliver cargo to destinations classified as either "short range" (SR) or "medium range" (MR).

If license = 2.1, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD and destination = SR or MR.

If an operator is classified 2.2, then he or she can:
1) operate vehicles classified as "liquid", and "medium duty" (LQ-MD), in addition to the vehicles and destinations for which a driver with a 1.1, 1.2, 1.3, or a 2.1 license is licensed.

If license = 2.2, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD and destination = SR or MR.

If an operator is classified 2.3, then he or she can:
1) operate vehicles classified as "liquid", and "heavy duty" (LQ-HD), and also to make deliveries using the same vehicles and to the same destinations as a driver with a 1.1, 1.2, 1.3, 2.1 or a 2.2 license.

If license = 2.3, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD and destination = SR or MR.

Hazardous and Long Range

If an operator is classified 3.1, then he or she can:

1) operate vehicles classified "**general purpose**" or "**liquid**" and either "**light duty**" (GP-LD and LQ-LD), "**medium duty**" (GP-MD and LQ-MD), or "**heavy duty**" (GP-HD and LQ-HD) plus vehicles which are classified "**hazardous**" and "**light duty**" (HZ-LD).

2) deliver cargo to destinations classified "**short range**" (SR), "**medium range**" (MR), or "**long range**" (LR).

If license = 3.1, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD and destination = SR or MR or LR.

If an operator is classified 3.2, then he or she can operate:

1) vehicles classified "**hazardous**" and "**medium duty**" (HZ-MD) and also to make deliveries using the same vehicles and to the same destinations that a driver with a 1.1, 1.2, 1.3, 2.1, 2.2, 2.3 or a 3.1 license.

If license = 3.2, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD or HZ-MD and destination = SR or MR or LR.

If an operator is classified 3.3, then he or she can operate:

1) vehicles classified "**hazardous**" and "**heavy duty**" (HZ-HD) and also to make deliveries using the same vehicles and to the same destinations that a driver with a 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.1, or a 3.2 license.

If license = 3.3, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD or HZ-MD or HZ-HD and destination = SR or MR or LR.

Appendix A (continued)

Below is a graphic representation of the license rules. Each box in this table contains all of the licenses eligible to deliver a particular cargo based on its cargo type, weight, destination, vehicle, and distance to destination. For example, a light duty, liquid cargo (LQ-LD) could be delivered to a short range (SR) destination by operators with license grades of 2.1, 2.2, 2.3, 3.1, 3.2, and 3.3. (For more explanation, ask the experimenter for help now.)

DIST-ANCE	VEHICLE TYPE								
	GP-LD	GP-MD	GP-HD	LQ-LD	LQ-MD	LQ-HD	HZ-LD	HZ-MD	HZ-HD
Short Range	1.1 1.2 1.3 2.1 2.2 2.3 3.1 3.2 3.3	1.2 1.3 2.1 2.2 2.3 3.1 3.2 3.3	1.3 2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.2 2.3 3.1 3.2 3.3	2.3 3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3
Medium Range	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.2 2.3 3.1 3.2 3.3	2.3 3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3
Long Range	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3

THE TASK

Let's examine how all this comes together in the task. The experiment will be divided into sessions of 3 blocks of 36 trials each. You may take breaks between trials or between blocks. For each trial you will be presented with the following information in one computer:

1) The name of the cargo to be delivered,
2) The weight of the cargo in kilograms (kg),
3) The name of the vehicle with which to deliver the cargo
4) The name of the destination to which the cargo is to be delivered.

This display is the 'Study Display.' Please study this information. Based on this information (and what you know about the structure and rules of the task) you must decide which operator (or operators) can make the delivery. As soon as you have formulated a set of possible operators who can perform the task, press the spacebar and you will be presented with a display containing the names of four operators. (The minimum number of possible operators for any delivery is three. Think about it.) There will always be four names to choose from. One, and only one, of these names will be the best answer according to the 'decision dispatching rule.' The number of operators capable of performing the task will vary from trial to trial. Examine the choices and make your decision as quickly as possible (without sacrificing accuracy). When you have made your decision press the key on the numeric keypad corresponding to the position of your choice in the display. While this task will be extremely challenging it's not quite as bad as it might seem; we have provided on-line help. The nature of this help will be described below.

As soon as you press the key to indicate your choice you will be given feedback about your answer. When you make a correct decision, the screen will display the word "CORRECT", give the response time in milliseconds, and the correct answer will blink on and off. When you make an incorect answer, the screen will display the word "INCORRECT" and the correct answer will be highlighted. After you have made your selection and have been given feedback, you may again access help to understand your mistake. Work quickly, but above all, work accurately.

The help system provides three categories of information: classes, rules, and names. The following pages show the information contained in the help screens. After you have examined all of the help, please ask the experimenter if you have questions about any part of the task.

Appendix A (continued)

HELP INFORMATION SCREEN

The Help Information Screen provides you with a map by which you can navigate your way through the system. You may access this screen by pressing the 'H' key. The experimenter will show you the keys. You can access any help screen at any time by pressing the key corresponding to the type of help you need. To leave any of the help screens and return to the Study Display press the escape key ('Esc') located in the top left corner of the key board.

WEIGHT CLASSES W-C	DISTANCE CLASSES D-C		HELP MENU H
DESTINATION CLASSES De-C	LICENSE CLASSES L-C	VEHICLE CLASSES V-C	CARGO CLASSES C-C
DESTINATION RULES De-R	LICENSE RULES L-R	VEHICLE RULES V-R	DECISION RULE D-R
DESTINATION NAMES D-N	OPERATOR NAMES O-N	VEHICLE NAMES V-N	CARGO NAMES C-N

The next several pages show you the information you get when you access the different types of help. Please read through each type of description of help, and take note of the type of information that is available.

HELP: CLASS INFORMATION

The following tables present the cargo, distance, destination, license, vehicle, and weight classes, followed by their abbreviations and/or ranges where appropriate.

View the distance class screen by pressing the key marked 'D-C.'

DISTANCE CLASSES

short	(S)	0 - 80	km
medium	(M)	81 - 320	km
long	(L)	321+	km

View the Cargo Classes screen by pressing the key marked 'C-C.'

CARGO CLASSES

general purpose (GP)
liquid (LQ)
hazardous (HZ)

View the Weight Classes screen by pressing the key marked 'W-C.'

WEIGHT CLASSES

light	(L)	0 - 1,500	kg
medium	(M)	1,501 - 10,000	kg
heavy	(H)	10,001+	kg

Press the key marked 'V-C' to view the Vehicle Classes.

VEHICLE CLASSES

general purpose, light duty (GP-LD)
general purpose, medium duty (GP-MD)
general purpose, heavy duty (GP-HD)

---

liquid, light duty (LQ-LD)
liquid, medium duty (LQ-MD)
liquid, heavy duty (LQ-HD)

---

hazardous, light duty (HZ-LD)
hazardous, medium duty (HZ-MD)
hazardous, heavy duty (HZ-HD)

Appendix A (continued)

Press the key marked 'De-C' to view the possible Destination Classes.

DESTINATION CLASSES

general purpose, short range (GP-SR)
general purpose, medium range (GP-MR)
general purpose, long range (GP-LR)

---

liquid, short range (LQ-SR)
liquid, medium range (LQ-MR)
liquid, long range (LQ-LR)

---

hazardous, short range (HZ-SR)
hazardous, medium range (HZ-MR)
hazardous, long range (HZ-LR)

Press the key marked 'L-C' to view the License Class information.

LICENSE CLASSES

1.1: general purpose, light duty, short range (GP-LD-SR)
1.2: general purpose, medium duty, short range (GP-MD-SR)
1.3: general purpose, heavy duty, short range (GP-HD-SR)

2.1: liquid, light duty, medium range (LQ-LD-MR)
2.2: liquid, medium duty, medium range (LQ-MD-MR)
2.3: liquid, heavy duty, medium range (LQ-HD-MR)

3.1: hazardous, light duty, long range (HZ-LD-LR)
3.2: hazardous, medium duty, long range (HZ-MD-LR)
3.3: hazardous, heavy duty, long range (HZ-HD-LR)

Appendix A (continued)

HELP: RULE INFORMATION

Some help screens provide information about the rules that regulate the performance of the task.

Access the rules governing Vehicles by pressing the 'V-R' key.

RULES GOVERNING VEHICLES
===== ========= ========

Any vehicle can travel any distance to deliver its cargo.

If vehicle = LD, then cargo weight ≤ 1,500 kg.

If vehicle = MD, then cargo weight ≤ 10,000 kg.

If vehicle = HD, then there is no maximum cargo weight.

If vehicle = GP, then cargo = GP.

If vehicle = LQ, then cargo = LQ.

If vehicle = HZ, then cargo = HZ.

Press the key marked 'De-R' to view the Destination rules.

RULES GOVERNING DESTINATIONS
===== ========= ============

Any destination can receive any amount (i.e., weight) of cargo.

If destination = GP, then cargo = GP.

If destination = LQ, then cargo = LQ.

If destination = HZ, then cargo = HZ.

If destination = SR, then distance ≤ 80 km.

If destination = MR, then 80 km < distance ≤ 320 km.

If destination = LR, then distance > 320 km.

Appendix A (continued)

LICENSE RULES

Press the 'L-R' key to view the rules governing Licenses.

DIST-ANCE	VEHICLE TYPE								
	GP-LD	GP-MD	GP-HD	LQ-LD	LQ-MD	LQ-HD	HZ-LD	HZ-MD	HZ-HD
Short Range	1.1 1.2 1.3 2.1 2.2 2.3 3.1 3.2 3.3	1.2 1.3 2.1 2.2 2.3 3.1 3.2 3.3	1.3 2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.2 2.3 3.1 3.2 3.3	2.3 3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3
Medium Range	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.1 2.2 2.3 3.1 3.2 3.3	2.2 2.3 3.1 3.2 3.3	2.3 3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3
Long Range	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.1 3.2 3.3	3.2 3.3	3.3

Press the 'D-R' key to view the Dispatching Decision Rule.

DISPATCHING DECISION RULE

The operator with the lowest license classification qualified to operate the available vehicle and qualified to deliver the cargo to the required destination is to be given the assignment.

HELP: NAME INFORMATION

The Name Help Information Screens show the Cargos, Vehicles, Destinations, or Operators that belong to each of the classes.

View the Cargo Names by pressing the key marked 'C-N.'

CARGO

GP	LQ	HZ
lumber	water	mercury
books	milk	cobalt
clothes	whisky	asbestos

View the Vehicle Names by pressing the key marked 'V-N.'

VEHICLES

GP-LD	GP-MD	GP-HD
Load Hog 1000	Load Hog 2000	Load Hog 3000
Freight King 100	Freight King 200	Freight King 300

LQ-LD	LQ-MD	LQ-HD
Tank King 1000	Tank King 2000	Tank King 3000
Route Master 100	Route Master 200	Route Master 300

HZ-LD	HZ-MD	HZ-HD
Haul Master 1000	Haul Master 2000	Haul Master 3000
Kargo King 100	Kargo King 200	Kargo King 300

View the Destination Names by pressing the key marked 'D-N.'

DESTINATIONS

GP-SR	GP-MR	GP-LR
United Enterprises	Olympia Industries	Island Enterprises
Keystone Systems	Matrix Co.	Universal Systems
Paragon Inc.	Globe Products	Standard Corp.

LQ-SR	LQ-MR	LQ-LR
National Systems	Horizon Technology	Victory Corp.
Republic Enterprises	Acme Corp.	Ajax Industries
Phoenix Technology	Fidelity Systems	Excel Services

HZ-SR	HZ-MR	HZ-LR
Charter Systems	Marathon Corp.	Colonial Inc.
Federal Assoc.	Western Enterprises	Vulcan Assoc.
Triad Co.	Heritage Ltd.	Beta Corp.

Appendix A (continued)

View the Operator Names by pressing the key marked 'O-N.'

OPERATORS

1.1: GP-LD-SR	1.2: GP-MD-SR	1.3: GP-HD-SR
Eloise	Bradley	Eugene
Julian	Agatha	Lester
Gwen	Conrad	Gina

2.1: LQ-LD-MR	2.2: LQ-MD-MR	2.3: LQ-HD-MR
Lolita	Valerie	Herbert
Rosalie	Vance	Vera
Barney	Mable	Adele

3.1: HZ-LD-LR	3.2: HZ-MD-LR	3.3: HZ-HD-LR
Nelson	Bernice	Enid
Felix	Troy	Vincent
Claude	Olivia	Stella

APPENDIX B. POST-EXPERIMENTAL QUESTIONNAIRE

Following Transfer Session

If you were giving advice to someone who was going to do the dispatcher task for the first time, what pointers or tips would you give them?

Condition 1

Subject 1. When I did the task, I confused the destination grouping with the operator grouping by thinking that all GP destinations were SR, etc. The weight of the cargo isn't important at all, just the size of the vehicle. I didn't memorize or look at the help menu for vehicles. I just looked at the number in the vehicle name, and if it began with a "1" it was LD, "2" was MD, etc. I looked at the cargo first, then vehicle, and then location just so I would have the information in the order I memorized the *operator licenses (for example:* GP-HD-LR).

Subject 3. I would tell them to be especially careful when looking at the weight of the cargo and not taking into consideration the vehicle, because the vehicle is what determines the driver and not the weight, which was a common mistake. I would also recommend that they make note of the fact that the combination of vehicle to location plays an important role in the decision-making process. Lastly, don't rush and select the name that is in the category, because that is not always the case.

Subject 4. Look first at the location. Then check the operator names to eliminate some of your choices. Then look at weight and cargo to further eliminate choices.

Subject 5.
1. Ignore the weight, the type of vehicle is the important information.

2. Determine what cargo you have and where the destination is, as well as what duty vehicle and then look at the licensing.
3. Then look at the names.

Subject 6. Tell them to memorize the names before the destinations-it is easier. You really need the names in order to answer the final question. You can check the destination before deciding which license number is best, but once you get to the end you must know the names in order to answer correctly. To remember the names and destinations, visualize the whole chart in your head, not just one group at a time. Say all group names in order in head. Easy to keep straight what goes where. Don't worry so much about time-it stresses you out.

Subject 7. First look at the destination-if it's light duty (sic) then the driver choice depends on the truck, 100, 200, 300. If it's MD, the driver must come from 2.0 range-then consider truck value. If it's HD, the driver must come from 3.0 range. Once you've established the license level the .1, .2, or .3 depends solely on the truck. Example: If the destination is marathon the 3.0 level is established, so if the truck Kargo King 100 is available, the best drivers would be 3.1: Nelson, Felix, or Claude.

Condition 2

Subject 13. Look at the vehicle, it will tell you what type of cargo you are carrying, also it will tell you the "duty" of the weight. Then look at the location to see if it is a SR, MR, or Long Range destination. Then take the information of weight (LD, MD, HD), cargo (GP, LQ, HZ), and location (SR, MR, LR) and compare it to the drivers' credentials. Memorize where the optimal position is on the screen and then look at the names given. Take the closest name to the optimal position.

Subject 14. Only look at location and weight. Figure out eligible drivers from that information (license classes menu).

Subject 15. Memorize names and destinations according to certain categories. For example, try to remember all the people who can drive hazardous loads or try to remember all the businesses with medium range destinations. Work carefully and develop some type of pattern. You will eventually be as accurate with much greater speed.

Subject 16. Make sure you know the names of the dispatchers and learn how to recognize the vehicles' names (i.e., 100 or 1000 LD, 200 or 2000 MD, and 300 or 3000 HD). That is more important than the weight itself. Then look at material so you can get GP LD, GP MD, GP HZ, or GP. Look at the table, then look at destination, then look at rules table. Make sure and know dispatchers's names.

Subject 17. Do not be discouraged by low scores. You will get the hang of it. Remember that half of the information is unnecessary.

1. Examine weights for LD, MD, or HD.
2. Go to location and see what class the cargo is in and the distance.
3. Go to the license classes and find the correct one. Choose the lowest number.
4. Remember HD can do lower weights than itself.
5. Ignore the specific distances and type of truck. This information is unnecessary.

Subject 19. First, look at cargo-if it is GP, start at row 1 of operators; if it is LQ, eliminate row 1 of operators and start at row 2; and if it is HZ, eliminate rows 1 and 2 and start at row 3. Next look at location. If it is SR, you can start at row 1; if it is MR, eliminate row 1 and start at row 2; if it is LR, eliminate rows 1 and 2 and start at row 3. The first step takes

priority. Example: If you have LQ cargo, you choose row 2 and eliminate row 1 even if it is SR destination. To further break it down, use the weight of cargo and vehicle type. LD is column of operators, MD is column 2, and HD is column 3. Steps 1 and 2 take priority. For example, if you have LQ cargo, you can start at 2.1. If you have SR, again start at 2.1. If you have MD, start at 2.2. Any driver with 2.2 or higher license can transport the cargo.

Condition 3

Subject 25. Don't bother looking at the cargo or weight class, because the truck must be able to carry that cargo and weight. It's a given. Look first at the vehicle to determine what class you need to be to drive it. Then look at the destination. Some destinations will require a higher classed driver, just because of the distance. Finally, look at the operator names and determine who would be the best first choice, etc. I tended to envision the tables in my head (especially for operators). I also tend to speak the names out loud for memorization. Good luck and have fun.

Subject 26. The vehicle names are easy to remember. The cargo already states which type they are (GP, HZ, LQ) and the number tells the weight class. Therefore, the only thing necessary to check is the destination and the operator names (keeping in mind destination and vehicle). Also, the destination need not be checked for hazardous materials, since at least a 3.1 will be required anyway.

Subject 27. Ignore weight and cargo, as this information is irrelevant. It can be determined or inferred by the vehicle type.

Subject 28.

- Accuracy before speed.

- Only vehicle name chart, operator name chart, and destination chart are important; the others you should know as part of the instructions needed.
- The order of importance is vehicle, location, weight, cargo-look at them in this order for a faster decision time.

Subject 29. If you look at the type of vehicle, then you will automatically know the weight and cargo.

Condition 4

Subject 37. Concentrate on vehicle class and location. Certain trucks caused certain cargo and certain distance, with any load. Drivers are based on type of vehicle rather than weight.

Subject 38.

A. Accumulate license conditions:

1. First get GP/LQ/HZ and LD/MD/HD from vehicle.

2. Then get distance from vehicle.

3. Locate first heading in operators' names that matches #1 above, then see if distance from #2 is greater than distance in that heading. If not, you have the right heading. If so, move to the first column that matches distance

B. Memorize next nine operator names from that point (A3) on; then do dispatch.

Subject 39. Look at the vehicle classification first. This will determine if you have to pay any attention to the specified weight. Associate things with each other. For example, associate cargos with destinations. This way, everything won't have to be completely memorized.

Subject 40. Don't look at weight or cargo.

1. First look at vehicle VN - gives class and LD, MD, or HD.

2. Look at destination name De-N - tells SR, LR, or MR.

3. Look at names under class, duty, and distance you obtained.

Subject 41.

* Look at the cargo before looking at company names because you will have a better idea which row to search (and if hazardous not to search) in determining the distance of the company.

* You don't need to look at the vehicle sheet, just notice the number next to the vehicle name.

* Try to memorize the lists of operators as soon as possible.

Subject 42. Read instructions carefully. Take your time reading them. Don't look for patterns or tricks in the order or in the information that was given to you. Try to realize what keys you really need to use and what you don't. Pay attention to what the keys stand for (learn the abbreviations). Develop a regular order or routine on how you pick the dispatcher. Try to do it the same way every time. You may discover you can skip a step or two that you are doing.

Following Retention

Condition 1

Subject 2. To know the relationships between the operator table and the destination table. You have to realize that the range increases down the operator table, but decreases across the destination table. Also, it is more useful to look at the vehicle name, instead of the cargo weight. The name itself isn't important, but if the number begins with a 1, which is followed by zeroes, it is LD, 2 is MD, etc.

Subject 3. I would suggest that the person use some form of grouping technique for the drivers and the destinations. This would allow one to process the information quicker and with more efficiency.

Subject 4. First look at information given. Then look for destination. Look at operator names. Look at cargo, and then look at operator names again. Look at vehicle names then look at the operator names again. Eliminate any groups that cannot drive the vehicle or who do not transport heavier cargo.

Subject 5.
1. Decide whether the cargo is general purpose, liquid, or hazardous.
2. Decide if the vehicle is light duty, medium duty, or heavy duty. This can be done by looking at the numbers after the vehicle names (i.e., 1000=LD, 2000=MD, etc.).
3. Decide if the destination is short, medium, or long range (this can be done by using the Help).
4. Look at the license chart and put all the pieces together. The chart will tell you what levels of operators are appropriate.
5. Look at the operators listed (all of them).

6. Discard the weight information, it is not important-the vehicle type is the only important information for weight.

Subject 6. Be sure to learn the names in order because they are not always brought up in the earlier sections. Sometimes the ones from the end are brought up, so you need to have all the names in your head. Don't worry so much about the destinations because you can look them up quickly. Don't even worry about the weight given in the first information. That is included in the vehicle-so it wastes time to worry about it.

Subject 7. Look first at the destination. If the destination is:

- short range then license class 1 will do.
- mid range then license class 2
- long range then license class 3

After deciding 1, 2, or 3, to get decimal .1, .2, or .3 look at truck. If truck is lower than chosen class .1 - .3 will do. If in same class, then the 100, 200, or 300 decides the decimal place. Don't worry about the cargo or weight.

Condition 2

Subject 13. Look at the destination-see what range it was in. Then look at the vehicle, this will tell you the type of cargo and the "duty" of the cargo. Combine the type (GP, LQ, HZ) with the duty (HD, MD, LD) and then that will give you a classification. Then look at the distance (SR, MR, LR) and pick the first one in that category that is past (or on) the earlier classification. Look up the operators in that block which meet all the requirements, then pick the first operator on or past that block.

Subject 14. Take note of location, cargo type and cargo location, then, along with the weight, figure out what lowest license class will work.

Subject 15. Ignore the weight, because the vehicle type determines who can or cannot drive that shipment. First look at the vehicle type and the cargo. They will tell you where to begin your search for appropriate drivers (by column). Then look at the destination. It will tell you which row to find the appropriate driver in. When you have singled out the least qualified for the haul, memorize the next three groups of names. Then go to your choices and choose the correct response.
Subject 16. Pay attention to what you are doing. Try to organize so you can get a pattern of doing things.

Subject 17. Look at the weight and decide if it is light, medium, or heavy. Then look at cargo and find out if it is general, liquid, or hazardous. Then look at license rules, and then operator names. Choose the operator with the lowest number. Basically, I ignored companies and vehicles. They did not seem to have much relevance.

Subject 19. Look at destination first. Is it SR, MR, LR? Is it general, liquid, or hazardous? If you can't remember which company is general, liquid, or hazardous, look at cargo type next. Look at vehicle. Is it LD, MD, or HD? Destination range is very important.
1. Look at destination first. Is it SR, MR (if so, ignore first row operators); LR (if so, ignore first and second row operators). Once you decide which row to start on,
2. Look at type of company: general, liquid, or hazardous.
If it is hazardous, must stay to last row of operators.
If it is liquid, must stay on second or third row.
3. Look at type of vehicle.

Condition 3

Subject 25. Memorization of cargo types, as well as weight and vehicles, comes rather quickly. With knowledge of these things

all that leaves is looking up the distance to destination and operators. Upon determining distance, look at the operator file and read over those which are possible candidates in order of importance.

Subject 26. Check location first and then vehicle. Checking the vehicle makes weight and cargo unnecessary information. Glance at names in license classes you are looking for before making the final decision.

Subject 27. Weight and cargo are irrelevant. You need only look at the distance and vehicle type to determine driver.

Subject 28. You should remember to check the qualifications of each class of driver; i.e., their distances, cargo load, and cargo type. Pick the driver that is the least qualified for the job, but one that is qualified. Also, the least qualified driver is not always one of the choices so you must mentally go through the classes to find the choice listed that is the least qualified for that particular task.

Subject 29. Look at vehicle type to get cargo and weight type. Look at destination for the range.

Subject 30. First look at the vehicle type-usually a number like 100 or 1000 will tell you if it is LD (1), MD (2), or HD (3). Then look at company, then go look at chart of all information and then try to memorize the first two or three name groups.

Condition 4

Subject 37. To concentrate on distance and destination. Use help frequently, especially the drivers' names to help remember them.

Subject 38. Once you understand the rules, the most useful help keys are vehicle names to get class and weight, then destination names to add distance to this; then start with that heading in operators' names and memorize that and the next two headings worth of operator names.

Subject 39.

* Look at the cargo first, then to the other items.
* Make sure you accurately know the list of drivers and their license specifications.
* Don't rush yourself.
* Be aware of what driver has preference.

Subject 40. Look at vehicle name to get type and duty. Look at destination to get range. Look at operators to see which can perform all three functions. Don't look at weights or cargo.

Subject 41. If you look at cargo before location you will know which row to look for it in. The location is not important in the hazardous materials group because any driver can go any distance. If the destination is farther than any driver in the cargo class can go, skip to the next row without bothering to look at weight or vehicle number. The first numerical character in the vehicle name is the same as the low (1), medium (2), or high (3) vehicle's capacity to carry its load. I always review operators before the test question to get my mind in order and review information.

Subject 42. Read instructions carefully. Develop a pattern of how you look at information. Learn your keys well. Look at information in same way every time. Take your time at first; it will save time later.

APPENDIX C. MISCELLANEOUS INDICES OF DISPATCHING-TASK PERFORMANCE

The tables contained within this appendix present indices of performance in the dispatching task that were not included in the text. When means and standard deviations are presented for a measure, the mean score is above the standard deviation; the standard deviation is enclosed in parentheses. All times are in seconds.

Appendix C (continued)

C-1. Mean Number of Keys Pressed.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				11.74 (06.16)
1	2				08.48 (02.28)
1	3				07.83 (08.03)
2	1			10.18 (05.64)	08.03 (02.23)
2	2			07.61 (02.74)	07.11 (02.32)
2	3			06.80 (02.49)	07.07 (02.38)
3	1				07.13 (02.53)
3	2				06.87 (02.50)
3	3				06.55 (02.48)
Transfer	1	06.49 (04.92)	08.71 (06.59)	06.22 (02.87)	06.66 (02.48)
Transfer	2	05.01 (02.43)	07.11 (03.59)	05.33 (02.12)	06.33 (02.37)
Transfer	3	04.72 (02.40)	06.57 (03.28)	05.40 (02.02)	06.46 (02.42)
Retention	1	07.74 (05.50)	08.65 (04.93)	07.68 (03.91)	09.19 (06.89)
Retention	2	06.05 (02.81)	07.03 (03.43)	06.76 (02.34)	07.13 (02.41)
Retention	3	05.94 (02.90)	06.26 (02.28)	06.49 (01.81)	06.85 (02.41)

Appendix C (continued)

C-2. Mean Study Time.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				12.98 (10.52)
1	2				08.02 (03.87)
1	3				07.40 (04.32)
2	1			12.76 (09.21)	07.54 (05.07)
2	2			08.69 (04.65)	06.18 (03.17)
2	3			06.75 (03.04)	05.81 (03.19)
3	1				05.46 (03.43)
3	2				04.76 (02.34)
3	3				04.48 (03.40)
Transfer	1	14.40 (10.56)	14.63 (10.22)	06.64 (04.29)	04.72 (02.92)
Transfer	2	09.95 (06.55)	08.60 (05.26)	05.50 (02.63)	04.25 (02.46)
Transfer	3	09.37 (07.47)	07.83 (06.73)	05.07 (02.42)	04.38 (03.12)
Retention	1	11.39 (09.08)	08.91 (07.58)	06.57 (04.68)	08.69 (09.20)
Retention	2	08.37 (06.43)	06.68 (03.68)	05.17 (02.67)	06.36 (05.66)
Retention	3	08.02 (06.27)	05.66 (03.09)	04.73 (02.40)	05.18 (03.31)

Appendix C (continued)

C-3. Mean Decision Latency.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				05.95 (06.10)
1	2				06.70 (06.66)
1	3				04.34 (03.96)
2	1			04.10 (02.90)	04.86 (04.68)
2	2			03.78 (02.51)	04.31 (03.22)
2	3			03.33 (02.31)	04.41 (04.57)
3	1				03.39 (03.65)
3	2				03.93 (04.85)
3	3				04.17 (04.71)
Transfer	1	04.78 (03.79)	05.43 (04.32)	02.54 (01.65)	04.65 (05.96)
Transfer	2	04.02 (03.12)	04.36 (02.98)	02.78 (01.97)	04.93 (07.16)
Transfer	3	03.98 (03.71)	04.35 (02.96)	02.87 (02.12)	04.82 (05.99)
Retention	1	04.22 (04.46)	04.13 (03.58)	02.76 (02.13)	04.70 (04.17)
Retention	2	03.97 (03.32)	04.17 (03.06)	02.98 (02.11)	05.04 (05.44)
Retention	3	03.78 (03.13)	04.54 (04.38)	03.19 (02.52)	05.22 (05.30)

Appendix C (continued)

C-4. Mean Time in All Help Screens, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				44.97 (37.36)
1	2				20.33 (16.47)
1	3				14.47 (12.00)
2	1			18.87 (23.45)	16.80 (13.01)
2	2			08.08 (07.64)	08.76 (06.57)
2	3			05.85 (05.69)	08.55 (06.62)
3	1				08.31 (07.72)
3	2				07.27 (06.49)
3	3				05.41 (05.63)
Transfer	1	08.96 (13.89)	12.79 (14.91)	04.25 (05.14)	05.94 (06.02)
Transfer	2	03.20 (05.04)	07.02 (06.79)	02.69 (03.77)	04.57 (04.65)
Transfer	3	02.50 (04.59)	05.37 (05.82)	02.51 (02.79)	04.27 (03.97)
Retention	1	12.75 (32.42)	12.57 (12.71)	07.27 (07.63)	14.43 (19.43)
Retention	2	04.75 (08.60)	06.84 (07.02)	04.96 (04.43)	07.20 (08.12)
Retention	3	03.22 (04.14)	05.57 (06.13)	04.34 (03.62)	06.13 (06.48)

Appendix C (continued)

C-5. Mean Time in "Classes" Help, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				02.34 (06.23)
1	2				00.25 (00.79)
1	3				00.10 (00.46)
2	1			00.45 (02.08)	00.05 (00.39)
2	2			00.00 (00.00)	00.00 (00.05)
2	3			00.00 (00.00)	00.00 (00.04)
3	1				00.00 (00.00)
3	2				00.02 (00.11)
3	3				00.00 (00.05)
Transfer	1	00.67 (03.10)	00.57 (02.07)	00.00 (00.00)	00.01 (00.11)
Transfer	2	00.04 (00.34)	00.04 (00.33)	00.00 (00.00)	00.00 (00.00)
Transfer	3	00.01 (00.09)	00.02 (00.25)	00.00 (00.00)	00.00 (00.08)
Retention	1	00.69 (03.89)	00.39 (02.27)	00.23 (01.34)	00.47 (02.73)
Retention	2	00.00 (00.00)	00.04 (00.42)	00.00 (00.00)	00.02 (00.17)
Retention	3	00.08 (00.51)	00.00 (00.00)	00.00 (00.06)	00.02 (00.16)

C-6. Mean Time in All Rules Help, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				02.28 (08.38)
1	2				00.66 (01.66)
1	3				00.78 (01.53)
2	1			02.52 (08.93)	01.26 (02.39)
2	2			00.65 (02.87)	01.06 (02.18)
2	3			00.28 (01.24)	00.88 (01.86)
3	1				00.86 (02.05)
3	2				00.94 (02.19)
3	3				00.51 (01.30)
Transfer	1	02.15 (05.96)	04.22 (08.00)	00.43 (01.65)	00.60 (01.41)
Transfer	2	01.00 (02.25)	01.99 (02.81)	00.49 (01.62)	00.67 (01.57)
Transfer	3	00.94 (02.50)	01.60 (02.19)	00.36 (01.15)	00.59 (01.26)
Retention	1	03.24 (12.64)	02.84 (05.22)	00.52 (01.85)	01.95 (04.35)
Retention	2	01.25 (02.49)	01.48 (02.66)	00.53 (01.87)	01.28 (05.17)
Retention	3	00.92 (01.71)	01.24 (02.04)	00.44 (01.41)	00.74 (01.61)

Appendix C (continued)

C-7. Mean Time in All Names Help, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				39.95 (29.24)
1	2				19.42 (16.15)
1	3				13.58 (11.77)
2	1			14.97 (12.78)	15.49 (12.32)
2	2			07.42 (06.33)	07.70 (06.06)
2	3			05.56 (05.00)	07.67 (06.33)
3	1				07.44 (06.96)
3	2				06.31 (05.63)
3	3				04.90 (05.17)
Transfer	1	06.06 (10.03)	07.12 (09.25)	03.82 (04.23)	05.32 (05.59)
Transfer	2	02.16 (04.81)	04.73 (06.51)	02.20 (02.67)	03.90 (04.12)
Transfer	3	01.55 (04.05)	03.55 (05.60)	02.15 (02.25)	03.66 (03.59)
Retention	1	08.34 (15.52)	09.02 (09.16)	06.49 (06.34)	11.79 (14.52)
Retention	2	03.49 (07.50)	05.33 (05.48)	04.43 (03.59)	05.90 (05.96)
Retention	3	02.22 (03.34)	04.33 (05.14)	03.90 (02.91)	05.37 (06.13)

C-8. Mean Time in Operator Names Help, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				30.01 (24.84)
1	2				14.88 (14.99)
1	3				09.56 (09.48)
2	1			07.93 (07.40)	11.23 (09.61)
2	2			03.81 (03.90)	04.91 (05.17)
2	3			03.22 (03.41)	04.84 (05.54)
3	1				04.69 (05.92)
3	2				03.76 (04.23)
3	3				02.66 (04.12)
Transfer	1	02.99 (07.08)	03.21 (06.47)	01.64 (02.35)	03.13 (04.51)
Transfer	2	01.33 (03.48)	03.22 (04.86)	01.08 (01.43)	01.94 (03.02)
Transfer	3	00.93 (02.81)	02.14 (04.65)	01.15 (01.28)	01.71 (02.51)
Retention	1	04.14 (06.96)	05.03 (07.18)	03.50 (03.41)	07.93 (11.21)
Retention	2	01.51 (03.60)	03.25 (04.34)	02.76 (02.71)	03.03 (04.52)
Retention	3	01.03 (01.96)	02.39 (04.59)	02.27 (01.52)	02.84 (05.00)

Appendix C (continued)

C-9. Mean Time in Destination Names Help, Pre-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				06.01 (04.88)
1	2				02.83 (02.17)
1	3				02.56 (02.19)
2	1			03.19 (04.10)	02.55 (02.66)
2	2			01.95 (02.30)	01.73 (01.69)
2	3			01.54 (02.05)	01.72 (01.56)
3	1				01.70 (01.61)
3	2				01.45 (01.57)
3	3				01.27 (01.34)
Transfer	1	01.38 (02.80)	01.89 (02.83)	01.30 (01.95)	01.28 (01.35)
Transfer	2	00.71 (01.87)	01.03 (01.90)	00.82 (01.55)	01.09 (01.24)
Transfer	3	00.60 (01.59)	01.00 (01.51)	00.86 (01.38)	01.14 (01.26)
Retention	1	02.67 (05.02)	02.58 (02.88)	02.18 (02.76)	02.39 (02.74)
Retention	2	01.25 (02.24)	01.92 (02.20)	01.43 (01.65)	02.13 (02.59)
Retention	3	00.96 (01.36)	01.83 (01.87)	01.52 (02.04)	01.77 (01.62)

Appendix C (continued)

C-10. Mean Time in All Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				06.67 (22.18)
1	2				01.46 (04.63)
1	3				00.63 (03.72)
2	1			03.18 (10.89)	00.45 (02.40)
2	2			01.64 (13.95)	00.14 (01.29)
2	3			00.33 (01.51)	00.46 (05.60)
3	1				00.07 (00.65)
3	2				00.05 (00.55)
3	3				00.04 (00.46)
Transfer	1	01.86 (09.90)	03.70 (16.09)	00.08 (00.64)	00.01 (00.18)
Transfer	2	01.39 (09.56)	01.17 (05.98)	00.05 (00.51)	00.01 (00.18)
Transfer	3	00.19 (01.62)	00.28 (01.71)	00.10 (01.38)	00.00 (00.00)
Retention	1	01.93 (07.74)	02.48 (12.89)	00.36 (02.37)	00.62 (05.69)
Retention	2	00.61 (02.59)	01.45 (05.72)	00.29 (01.58)	00.54 (07.08)
Retention	3	00.34 (01.95)	00.08 (00.87)	00.10 (00.56)	00.00 (00.00)

Appendix C (continued)

C-11. Mean Time in Classes Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				00.02 (00.24)
1	2				00.00 (00.00)
1	3				00.00 (00.00)
2	1			00.24 (02.57)	00.00 (00.00)
2	2			00.92 (13.53)	00.00 (00.00)
2	3			00.00 (00.00)	00.09 (01.27)
3	1				00.00 (00.00)
3	2				00.00 (00.00)
3	3				00.00 (00.00)
Transfer	1	00.51 (04.76)	00.46 (02.70)	00.00 (00.00)	00.00 (00.00)
Transfer	2	00.00 (00.03)	00.09 (01.26)	00.00 (00.00)	00.00 (00.00)
Transfer	3	00.00 (00.00)	00.00 (00.00)	00.06 (00.84)	00.00 (00.00)
Retention	1	00.04 (00.55)	00.12 (01.62)	00.00 (00.00)	00.02 (00.31)
Retention	2	00.00 (00.00)	00.03 (00.40)	00.01 (00.11)	00.00 (00.00)
Retention	3	00.10 (00.82)	00.00 (00.00)	00.00 (00.00)	00.00 (00.00)

C-12. Mean Time in Rules Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				00.41 (04.01)
1	2				00.01 (01.16)
1	3				00.00 (00.00)
2	1			00.28 (02.57)	00.00 (00.00)
2	2			00.00 (00.00)	00.00 (00.00)
2	3			00.00 (00.00)	00.25 (03.63)
3	1				00.00 (00.00)
3	2				00.00 (00.00)
3	3				00.00 (00.00)
Transfer	1	00.78 (04.90)	00.80 (04.49)	00.00 (00.07)	00.00 (00.00)
Transfer	2	00.42 (02.39)	00.16 (01.39)	00.00 (00.00)	00.00 (00.00)
Transfer	3	00.03 (00.40)	00.00 (00.00)	00.00 (00.00)	00.00 (00.00)
Retention	1	00.38 (03.69)	00.63 (06.78)	00.00 (00.07)	00.37 (03.18)
Retention	2	00.00 (00.00)	00.06 (00.59)	00.00 (00.00)	00.48 (07.03)
Retention	3	00.02 (00.30)	00.00 (00.00)	00.00 (00.00)	00.00 (00.00)

Appendix C (continued)

C-13. Mean Time in Names Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				05.61 (19.77)
1	2				01.42 (04.60)
1	3				00.63 (03.72)
2	1			02.46 (07.57)	00.42 (02.35)
2	2			00.72 (03.57)	00.13 (01.26)
2	3			00.33 (01.51)	00.12 (00.90)
3	1				00.06 (00.60)
3	2				00.04 (00.44)
3	3				00.04 (00.46)
Transfer	1	00.39 (02.34)	01.96 (11.04)	00.07 (00.51)	00.01 (00.18)
Transfer	2	00.98 (08.03)	00.85 (04.12)	00.05 (00.51)	00.01 (00.18)
Transfer	3	00.09 (00.83)	00.27 (01.70)	00.05 (00.55)	00.00 (00.00)
Retention	1	01.38 (04.64)	01.67 (05.52)	00.33 (02.31)	00.17 (02.27)
Retention	2	00.59 (02.50)	01.30 (05.21)	00.27 (01.55)	00.04 (00.64)
Retention	3	00.19 (01.00)	00.08 (00.87)	00.10 (00.56)	00.00 (00.00)

Appendix C (continued)

C-14. Mean Time in Operator Names Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				05.27 (18.91)
1	2				01.40 (04.60)
1	3				00.63 (03.72)
2	1			02.28 (06.69)	00.41 (02.31)
2	2			00.72 (03.57)	00.13 (01.26)
2	3			00.33 (01.50)	00.12 (00.90)
3	1				00.06 (00.60)
3	2				00.04 (00.44)
3	3				00.04 (00.46)
Transfer	1	00.29 (02.04)	01.28 (06.68)	00.06 (00.49)	00.01 (00.18)
Transfer	2	00.34 (02.93)	00.70 (03.33)	00.05 (00.51)	00.01 (00.18)
Transfer	3	00.05 (00.54)	00.27 (01.70)	00.05 (00.55)	00.00 (00.00)
Retention	1	01.16 (03.94)	01.39 (04.10)	00.33 (02.31)	00.06 (00.80)
Retention	2	00.45 (01.56)	01.30 (05.21)	00.27 (01.54)	00.04 (00.64)
Retention	3	00.19 (01.00)	00.08 (00.87)	00.10 (00.53)	00.00 (00.00)

C-15. Mean Time in Destination Names Help, Post-Response.

Session	Block	Instruct. First	Instruct. Last	Alternat.	Task Only
1	1				00.30 (02.11)
1	2				00.00 (00.00)
1	3				00.00 (00.00)
2	1			00.08 (01.07)	00.00 (00.00)
2	2			00.00 (00.00)	00.00 (00.00)
2	3			00.01 (00.08)	00.00 (00.00)
3	1				00.00 (00.00)
3	2				00.00 (00.00)
3	3				00.00 (00.00)
Transfer	1	00.03 (00.31)	00.51 (04.13)	00.00 (00.07)	00.00 (00.00)
Transfer	2	00.63 (05.61)	00.10 (00.78)	00.00 (00.00)	00.00 (00.00)
Transfer	3	00.03 (00.34)	00.00 (00.00)	00.00 (00.00)	00.00 (00.00)
Retention	1	00.14 (01.07)	00.17 (01.47)	00.00 (00.00)	00.05 (00.63)
Retention	2	00.10 (00.91)	00.00 (00.02)	00.00 (00.05)	00.00 (00.00)
Retention	3	00.00 (00.00)	00.00 (00.00)	00.01 (00.08)	00.00 (00.00)

Appendix C (continued)

C-16. Number of Trials for Which a Particular Help Screen Was Accessed in Block 1, Transfer Session.

Training Group	Help Screen		
	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	4	1	4
Instr. Last	3	3	61
Alternating	-	-	-
Task Only	-	-	-
	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	14	3	3
Instr. Last	25	6	6
Alternating	-	2	-
Task Only	2	-	-
	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	5	7	61
Instr. Last	7	8	131
Alternating	-	-	104
Task Only	-	-	144
	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	4	7	72
Instr. Last	6	13	84
Alternating	-	1	81
Task Only	-	-	103
	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	7	39	66
Instr. Last	5	74	72
Alternating	-	16	134
Task Only	-	41	132

[a]The term Instr. is an abbreviation for the word instructions.

Appendix C (continued)

C-17. Number of Trials for Which a Particular Help Screen Was Accessed in Block 2, Transfer Session.

Training Group	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	-	1	-
Instr. Last	-	1	-
Alternating	-	-	61
Task Only	-	-	-

Training Group	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	2	-	1
Instr. Last	3	-	-
Alternating	-	-	-
Task Only	-	-	-

Training Group	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	1	-	44
Instr. Last	-	3	85
Alternating	-	-	75
Task Only	-	-	132

Training Group	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	2	2	21
Instr. Last	-	-	38
Alternating	-	-	29
Task Only	-	-	105

Training Group	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	2	48	52
Instr. Last	-	85	99
Alternating	-	21	130
Task Only	-	46	119

[a]The term Instr. is an abbreviation for the word instructions.

Appendix C (continued)

C-18. Number of Trials for Which a Particular Help Screen Was Accessed in Block 3, Transfer Session.

Training Group	Help Screen		
	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	-	-	-
Instr. Last	-	-	63
Alternating	-	-	-
Task Only	-	-	-
	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	1	-	-
Instr. Last	1	-	-
Alternating	-	-	-
Task Only	2	-	-
	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	-	-	43
Instr. Last	-	1	95
Alternating	-	-	88
Task Only	-	-	142
	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	-	-	7
Instr. Last	-	-	35
Alternating	-	-	12
Task Only	-	-	105
	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	1	57	46
Instr. Last	-	90	71
Alternating	-	21	136
Task Only	-	48	119

[a]The term Instr. is an abbreviation for the word instructions.

Appendix C (continued)

C-19. Number of Trials for Which a Particular Help Screen Was Accessed in Block 1, Retention Session.

Training Group	Help Screen		
	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	4	6	5
Instr. Last	5	3	19
Alternating	2	2	2
Task Only	6	5	6
	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	11	9	6
Instr. Last	13	5	3
Alternating	7	6	1
Task Only	17	8	7
	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	9	6	131
Instr. Last	3	11	148
Alternating	3	3	149
Task Only	5	8	182
	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	9	9	42
Instr. Last	4	10	63
Alternating	1	3	63
Task Only	7	7	98
	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	11	62	131
Instr. Last	3	74	144
Alternating	4	20	186
Task Only	6	67	187

[a]The term Instr. is an abbreviation for the word instructions.

Appendix C (continued)

C-20. Number of Trials for Which a Particular Help Screen Was Accessed in Block 2, Retention Session.

Training Group	Help Screen		
	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	-	-	-
Instr. Last	1	1	-
Alternating	-	-	-
Task Only	-	-	-
	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	-	-	-
Instr. Last	1	2	-
Alternating	-	-	-
Task Only	3	-	-
	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	-	-	109
Instr. Last	1	1	139
Alternating	-	1	137
Task Only	1	2	175
	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	-	1	25
Instr. Last	1	2	19
Alternating	-	-	36
Task Only	-	2	76
	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	-	66	89
Instr. Last	1	80	137
Alternating	-	21	189
Task Only	1	54	156

[a]The term Instr. is an abbreviation for the word instructions.

Appendix C (continued)

C-21. Number of Trials for Which a Particular Help Screen Was Accessed in Block 3, Retention Session.

Training Group	Help Screen		
	DECISION RULE	DISTANCE CLASSES	HELP MENU
Instr.[a] First	1	-	-
Instr. Last	-	-	-
Alternating	-	-	-
Task Only	-	-	-
	WEIGHT CLASSES	CARGO NAMES	CARGO CLASSES
Instr. First	-	-	1
Instr. Last	-	-	-
Alternating	1	1	-
Task Only	3	-	-
	DESTINATION CLASSES	DESTINATION RULES	DESTINATION NAMES
Instr. First	1	-	112
Instr. Last	-	1	146
Alternating	-	-	140
Task Only	-	-	174
	VEHICLE CLASSES	VEHICLE RULES	VEHICLE NAMES
Instr. First	-	1	23
Instr. Last	-	4	12
Alternating	-	-	12
Task Only	-	-	70
	LICENSE CLASSES	LICENSE RULES	OPERATOR NAMES
Instr. First	6	61	85
Instr. Last	-	77	121
Alternating	-	22	194
Task Only	-	49	136

[a]The term Instr. is an abbreviation for the word instructions.

☆U.S. Government Printing Office: 1992 – 750-061/60231